LECTURE 2: INTRODUCTION 2

- During the last lecture we reviewed the basic toolkit of probability and Statistics, especially parts that find common use in physics.
- Today we delve further ~~into~~ and explore a few of these tools in further depth.
- The material we covered in the previous lecture normally finds use in Equilibrium Statistical Mechanics and in the linear response regime when systems are only nominally removed from thermal equilibrium.
- The concepts we explore today extend to the nonlinear regimes when physical systems are strongly driven and removed far from thermal equilibrium; the regime where Non-equilibrium Statistical Mechanics operates, and a sound theory of which has yet to be formulated.

* MOMENTS & CHARACTERISTIC FUNCTION:

- In the previous lecture we cursorily defined the moments of a distribution and quickly moved on to use the first two moments, viz. the mean $\langle x \rangle$ and variance $\langle x^2 \rangle$ or std. deviation σ.

- More generally, one can define higher-order moments of the distribution $p(x)$ as the average of powers of x:

$$m_n \equiv \langle x^n \rangle = \int x^n p(x)\, dx \longrightarrow (1)$$

- Accordingly, the mean m_1 is the first moment ($n=1$), and the variance is related to the second moment ($\sigma^2 = m_2 - m_1^2$)
- The above definition is only meaningful if the integral converges, which requires that $p(x)$ decreases sufficiently

rapidly for large $|x|$.

- From a theoretical point of view, the moments are interesting: if they exist, their knowledge is often equivalent to the knowledge of the distribution. But this statement is not rigorously correct, it holds most of the time, but there exist examples of different distributions densities which possess exactly the same moments.
- In practice however, the high order moments are very hard to determine satisfactorily: as n grows, longer and longer time-series are needed to keep a certain level of precision on m_n; these high moments are thus ~~not~~ in general not adapted to describe empirical data.

- For many computational purposes, it is convenient to introduce the Characteristic function of $p(x)$, defined as its Fourier transform:

$$\hat{p}(z) \equiv \int e^{izx} p(x)\, dx \longrightarrow (2)$$

- The function $p(x)$ is itself related to its characteristic function through an inverse Fourier transform:

$$p(x) = \frac{1}{2\pi} \int e^{-izx} \hat{p}(z)\, dz \longrightarrow (3)$$

- Since $p(x)$ is normalized, one always has $\hat{p}(0) = 1$.
- The moments of $p(x)$ can then be obtained through successive derivatives of the characteristic function at ~~$\hat{p}(0)$~~ $z=0$:

$$m_n = (-i)^n \frac{d^n}{dz^n} \hat{p}(z) \Big|_{z=0} \longrightarrow (4)$$

- One finally defines the <u>cumulants c_n</u> of a distribution as the successive derivatives of the logarithm of its characteristic function:

$$c_n = (-i)^n \frac{d^n}{dz^n} \log \hat{p}(z)\Big|_{z=0} \longrightarrow (5)$$

- The cumulant c_n is a polynomial combination of the moments m_p with $p \leq n$.
- For example $c_2 = m_2 - m_1^2 = \sigma^2$.
- It is often useful to normalize the cumulants by an appropriate power of the variance, such that the resulting quantities are dimensionless.
- One thus defines the normalized cumulants λ_n,

$$\lambda_n \equiv c_n / \sigma^n \longrightarrow (6)$$

- Often one uses the third & fourth normalized cumulants, called the <u>skewness</u> and <u>kurtosis</u> (κ):

$$\lambda_3 = \frac{\langle (x - \langle x \rangle)^3 \rangle}{\sigma^3} \; ; \; \lambda_4 = \frac{\langle (x - \langle x \rangle)^4 \rangle}{\sigma^4} \longrightarrow (7)$$

- In the previous lecture when we discussed stationary process in connection with the Autocorrelation function and power spectrum, we did not get into what constitutes a stationary time series.

- A process or time-series is considered stationary if all the moments of its distribution become time-invariant.
- But given ~~as~~ often finding high order moments is not possible, a weak requirement is the time-invariant convergence of the first four moments (or cumulants).

- The skewness quantifies the degree of symmetry in the distribution, whereas kurtosis quantifies its flatness.

- The above definition of cumulants may look arbitrary, but these quantities have remarkable properties.
- For example, the cumulants simply add when one sums independent random variables, a fact used in deriving the Central Limit Theorem.
- Furthermore, a Gaussian distribution is characterized by the fact that all cumulants of order larger than two are identically zero.
- Hence the cumulants, in particular κ, can be interpreted as a measure of the distance ~~from~~ between any given distribution $p(x)$ and a Gaussian.

*<u>DIVERGENCE OF MOMENTS - ASYMPTOTIC BEHAVIOR</u>: The moments (or cumulants) of a distribution do not always exist.
- A necessary condition for the n^{th} moment (m_n) to exist is that the distribution density $p(x)$ should decay faster than $1/|x|^{n+1}$ for $|x| \to \infty$, or else the integral in equation 1 would diverge for large $|x|$ values.
- If one only considers distribution densities that are behaving asymptotically as a power-law with an exponent ~~1+~~ $1+\xi$,

$$\cancel{p(x) \sim A_\pm}$$

$$p(x) \sim \frac{\xi A_\pm^{\xi}}{|x|^{1+\xi}} \text{ for } x \to \pm\infty \longrightarrow (8)$$

then all moments such that $n \geq \xi$ are infinite.

- For example, such a distribution has no finite variance whenever $\xi \leq 2$.
- Note that, for $p(x)$ to be a normalizable probability distribution, the integral $\int p(x)dx = 1$ must converge, this requires that $\xi > 0$.

- The characteristic function of a distribution having an asymptotic power-law behavior given by equation 8 is non-analytic around $z=0$.
- The small-z expansion contains regular terms of the form z^n for $n < \xi$ followed by a non-analytic term $|z|^\xi$ (possibly with logarithmic corrections such as $|z|^\xi \log z$, for integer ξ.
- The derivatives of order larger or equal to ξ of the characteristic function, thus does not exist at the origin ($z=0$).

× LÉVY DISTRIBUTIONS & PARETIAN TAILS: This is an ideal point to introduce another class of distributions in addition to ones we covered in the previous lecture, since we have now broached the topic of asymptotic power-law decay of distributions.

- Lévy distributions ($L_\xi(x)$) appear naturally in the context of the Central Limit Theorem because of their stability property under addition (a property also shared by Gaussians).
- The tails of Lévy distributions are however much 'fatter' than those of Gaussians, and are thus useful to describe multiscale phenomena, i.e. when both very large and very small values of a quantity can commonly be observed eg, amplitude of earthquakes or other natural catastrophes, ~~fluctuations~~ personal income etc.

- These distributions were introduced in the 1950s and 1960s by Mandelbrot (following Pareto) to describe personal income and the price of changes of some financial assets, in particular the price of cotton.
- An important constitutive property of these Lévy distributions is their power-law behavior for large arguments, often called 'Pareto tails'.

$$L_{\xi}(x) \sim \frac{\xi A_{\pm}^{\xi}}{|x|^{1+\xi}} \text{ for } x \longrightarrow \pm\infty \longrightarrow (9)$$

where $0 < \xi < 2$ is a certain exponent (often called α), and $A_{\pm}^{\xi}$ two constants which we call <u>tail amplitudes</u>, or <u>scale parameters</u>: A_{+} gives the order of magnitude of the large (positive or negative) fluctuations for x.
- For instance, the probability to draw a number larger than x decreases as $P_{>}(x) = (A_{+}/x)^{\xi}$ for large positive x.

- One can of course in principle observe Pareto tails with $\xi \geq 2$; but, those tails do not correspond to the asymptotic behavior of a Lévy distribution.

- In full generalization, Lévy distributions are characterised by an <u>asymmetry parameter</u> defined as $\beta \equiv (A_{+}^{\xi} - A_{-}^{\xi})/(A_{+}^{\xi} + A_{-}^{\xi})$, which measures the relative weight of the positive and negative tails.
- We shall mostly focus on the symmetric case $\beta = 0$.
- The fully asymmetric case $(\beta = 0)$ is also useful to describe strictly positive random variables, e.g. frequency of earthquakes of a given magnitude.

- An important consequence of equation 8 with $\xi_\varsigma \leq 2$ is that the variance of a Lévy distribution is formally infinite: the probability density does not decay fast enough for the integral $\sigma^2 = \int (x-\langle x\rangle)^2 p(x)\,dx$ to converge.
- In the case $\xi_\varsigma \leq 1$, the distribution density decays so slowly that even the mean fails to exist.
- What this means in practical terms is that if one measures a mean (for $\xi_\varsigma \leq 1$) or variance (for $\xi_\varsigma \leq 2$), one will certainly obtain a value, but it never converges but rather continues to drift with ever increasing samples added to the dataset.
- The scale of the fluctuations, defined by the width of the distribution, is always set by $A = A_+ = A_-$.

- There is unfortunately no simple analytical expression for the symmetric Lévy distributions $L_{\xi_\varsigma}(x)$, except for $\xi_\varsigma = 1$, which corresponds to a Cauchy distribution (or 'Lorentzian'): ~~we saw~~

$$L_1(x) = \frac{A}{x^2 + \pi^2 A^2} \longrightarrow (10)$$

- However, the characteristic function of a symmetric Lévy distribution is rather simple, it reads:

$$\hat{L}_{\xi_\varsigma}(z) = \exp\left(-a_{\xi_\varsigma} |z|^{\xi_\varsigma}\right) \longrightarrow (11)$$

where a_{ξ_ς} is a certain constant, proportional to the tail parameter A^{ξ_ς}; for example when $1 < \xi_\varsigma < 2$, $A^{\xi_\varsigma} = \xi_\varsigma \Gamma(\xi_\varsigma - 1) \sin(\pi \xi_\varsigma / 2) a_{\xi_\varsigma} / \pi$.
- It is thus clear that in the limit $\xi_\varsigma = 2$, one recovers the definition of a Gaussian.
- When ξ_ς decreases from 2, the distribution becomes more and more sharply peaked around the origin and fatter in its tails, while

'intermediate' events lose weight.

- These distributions therefore describe 'intermittent' phenomena, very often small, but sometimes gigantic fluctuations.

- Note finally that equation 11 does not define a probability distribution when $\xi > 2$, because it's inverse Fourier transform is not everywhere positive.
- In case $\beta \neq 0$, one would have:

$$\hat{L}^{\beta}_{\xi}(x) = \exp\left[-a_{\xi}|z|^{\xi}\left(1 + i\beta \tan(\xi\pi/2)\frac{z}{|z|}\right)\right] \quad (\xi \neq 1) \rightarrow (12)$$

- It is important to notice that while the leading asymptotic term for large x is given by equation 9, there are subleading terms which can be important for finite x.
- The full asymptotic series actually reads:

$$L_{\xi}(x) = \sum_{n=1}^{\infty} \frac{(-)^{n+1}}{\pi n!} \frac{a_{\xi}^{n}}{x^{1+n\xi}} \Gamma(1+n\xi) \sin(\pi\xi n/2) \rightarrow (13)$$

- The presence of the subleading terms may lead to a bad empirical estimate of the exponent ξ based on a fit of the tail of the distribution.

- In particular, the 'apparent' exponent which describes the function L_{ξ} for finite x is larger than ξ, and decreases towards ξ for $x \to \infty$, but more and more slowly as ξ gets nearer to the Gaussian value $\xi = 2$, for which the power-law tails no longer exist.
- Note however that one also often observes empirically the opposite behavior, i.e. an apparent Pareto exponent which <u>grows</u> with x.
- This arises when the Pareto distribution (equation 9), is only valid in an intermediate regime $x \ll 1/\alpha$, beyond which the distribution decays exponentially, say as $\exp(-\alpha x)$.
- The Pareto tail is then 'truncated' for large values of x, and this leads to an effective ξ which grows with x.

- An interesting generalization of Lévy distributions which accounts for this exponential cut-off is given by the 'Truncated Lévy distributions'.
- A simple way to alter the characteristic function (equation 11) to account for an exponential cut-off for large arguments is to set:

$$\hat{L}_{\xi}(z) = \exp\left[-a_{\xi}\frac{(\alpha^2+z^2)^{\frac{\xi}{2}}\cos\left(\xi \arctan(|z|/\alpha)\right) - \alpha^{\xi}}{\cos(\pi\xi/2)}\right] \longrightarrow (14)$$

for $1 \le \xi \le 2$.
- The above form reduces to eq^n 11 for $\alpha = 0$.
- Note that the argument in the exponential can also be written as

$$\frac{a_{\xi}}{2\cos(\pi\xi/2)}\left[(\alpha+iz)^{\xi} + (\alpha-iz)^{\xi} - 2\alpha^{\xi}\right] \longrightarrow (15)$$

Exponential Tail: A Limiting Case: Very often, it is seen that in the formal limit $\xi \to \infty$, the power-law tail becomes an exponential tail, if the tail parameter is simultaneously scaled as

$A^{\xi} = (\xi/\alpha)^{\xi}$

- ~~Chech~~ Qualitatively, this can be understood as follows: consider a probability distribution restricted to positive x, which decays as a power-law for large x, defined as:

$$P_>(x) = \frac{A^{\xi}}{(A+x)^{\xi}} \longrightarrow (16)$$

- This shape is obviously compatible with equation 9, and is such that $P_>(x=0) = 1$.
- If $A = (\xi/\alpha)$, one then finds:

$$P_>(x) = \frac{1}{[1+(\alpha x/\xi)]^{\xi}} \xrightarrow[\xi \to \infty]{} \exp(-\alpha x) \longrightarrow (17)$$

- From a physical standpoint, the exponential truncation of the power-law comes about due to finiteness of the physical system in which the power-law behavior is observed.
- For example, the ~~[illegible]~~ clouds in the atmosphere are known to possess a power-law distribution out to a distance of 1000 km.
- Beyond that point, the earth's curvature exerts itself and forces a cut-off in the form of an exponential decay.
- The above form is therefore useful in performing finite size scaling.
- Suppose that one could construct earths of various radii then the power-law size distribution of clouds would exhibit exponential cut-offs at various lengths, all of which can be re-scaled via the above arguments.

* EXTREME VALUE STATISTICS: Let us consider the statistics of academia.
- Starting from High School and Undergraduate study through graduate school and faculty position, we expect the top few percent of the candidate pool to proceed to the next stage.
- If the system of education is particularly competitive, let's imagine the top 1% of High School class move to undergraduate study, and thence the top ~~\$~~ 1% of the graduating undergraduate class is admitted to graduate school and so on.
- What statistics determine such a talent pool of students?
- Clearly this is not a Gaussian distributed population, for we are always selecting candidates from the high-performing tail.
- This is an exemplar of a class of problems where we are interested in considering a distribution built from extreme values.

- If one observes a series of N independent realizations of the same random phenomenon, the question we are asking is to determine the order of magnitude of the maximum observed value (or minimum, let's focus on maximum, the procedure is the same).

-

- The law of large numbers tells us that an event which has a probability p of occurrence appears on average Np times on a series of N observations.
- One thus expects to observe events which have a probability of at least $1/N$.
- It would be surprising to encounter an event with probability much smaller than $1/N$.

- The order of magnitude of the largest event, Λ_{max}, observed in a series of N independent identically distributed (iid) random variables is thus given by:

$$P_>(\Lambda_{max}) = 1/N \longrightarrow (18)$$

- More precisely, the full probability distribution of the maximum value $x_{max} = \max_{i=N}\{x_i\}$, is relatively easy to characterize; this will justify the above simple criterion in equation 18.
- The cumulative distribution $P(x_{max} < \Lambda)$ is obtained by noticing that if the maximum of all x_i's is smaller than Λ, then all of the x_i's must be smaller than Λ.
- If the random variables are iid, one finds:

$$P(x_{max} < \Lambda) = [P_<(\Lambda)]^N \longrightarrow (19)$$

- Note that this result is general, and does not rely on a specific choice for $p(x)$.
- When Λ is large, it is useful to use the following approximation:

$$P(x_{max} < \Lambda) = [P_<(\Lambda)]^N \longrightarrow (20)$$

- Since we now have a simple formula for the distribution of x_{max}, one can invert it in order to obtain, for example, the median value of the maximum, noted Λ_{med} such that $P(x_{max} < \Lambda_{med}) = 1/2$:

$$P_>(\Lambda_{med}) = 1 - (1/2)^{1/N} \simeq \frac{\log 2}{N} \longrightarrow (21)$$

- More generally, the value Λ_p which is greater than x_{max} with probability p is given by

$$P_>(\Lambda_p) \simeq -\frac{\log p}{N} \longrightarrow (22)$$

- The quantity Λ_{max} defined by equation 18 is thus such that $p = 1/e \simeq 0.37$.
- The probability that ~~Pmax~~ x_{max} is even larger than Λ_{max} is thus 63%.
- As we shall now see, Λ_{max} also corresponds, in many cases, to the most probable value of x_{max}.

~~Equation 18~~ Interestingly, the distribution of x_{max} only depends, when N is large, on the asymptotic behavior of the distribution of x, $p(x)$, when $x \to \infty$.

- For example, if $p(x)$ behaves as an exponential when $x \to \infty$, or more precisely if $P_>(x) \sim \exp(-\alpha x)$, one finds:

$$\Lambda_{max} = \frac{\log N}{\alpha} \longrightarrow 23$$

which grows very slowly with N. For instance, for a symmetric exponential distribution $p(x) = \exp(-|x|/2)$, the median value of the maximum of $N = 10^4$ variables is only 6.3.

- Setting $x_{max} = \Lambda_{max} + (u/\alpha)$, one finds that the deviation u around Λ_{max} is distributed according to the Gumbel distribution:

$$p(u) = e^{-e^{-x}} e^{-u} \longrightarrow 24$$

- The most probable value of this distribution is $u = 0$.
- This shows that Λ_{max} is the most probable value of x_{max}.
- The result (equation 24) is actually much more general, and is valid as soon as $p(x)$ decreases more rapidly than any power-law for $x \to \infty$: the deviation between Λ_{max} (as defined in equation 18) and x_{max} is always distributed according to the Gumbel law (equation 24) up to a scaling factor in defn of u.

- The situation is radically different if $p(x)$ decreases as a power-law.
- In this case,

$$\mathcal{P}_{>}(x) \simeq \frac{A_+^{\xi}}{x^{\xi}} \longrightarrow (25)$$

and the typical value of the maximum is given by:

$$\Lambda_{max} = A_+ N^{1/\xi} \longrightarrow (26)$$

- Numerically, for a distribution with $\xi = \frac{3}{2}$ and a scale factor $A_+ = 1$, the largest of $N = 10^4$ variables is of the order of 450, whereas for $\xi = 1/2$ it is 100 million!
- The complete distribution of the maximum, called the Fréchet distribution is given by:

$$p(u) = \frac{\xi}{u^{1+\xi}} e^{-1/u^{\xi}}; \quad u = \frac{x_{max}}{A_+ N^{1/\xi}} \longrightarrow (27)$$

- It's asymptotic behavior for $u \to \infty$ is still a power-law of exponent $1+\xi$.
- Said differently, both power-law tails and exponential tails are stable with respect to the 'max' operation.
- The most probable value x_{max} is now equal to $(\xi/(1+\xi))^{1/\xi} \Lambda_{max}$.
- As mentioned above, the limit $\xi \to \infty$ formally corresponds to an exponential distribution.
- In this limit, one indeed recovers Λ_{max} as the most probable value.

- Equation 26 allows us to intuitively discuss the divergence of the mean value for $\xi \leq 1$ and of the variance for $\xi \leq 2$.
- If the mean value exists, the sum of N random variables is typically equal to Nm_{ξ}, where m is the mean.
- But when $\xi < 1$, the largest encountered value of x is on the order of $N^{1/\xi} \gg N$, and would thus be larger than the entire sum.

- Similarly, when the variance exists, the RMS of the sum is equal to $\sigma\sqrt{N}$, but for $\xi < 2$, x_{max} grows faster than $\sqrt{N}$.

- More generally, one can rank the random variables x_i in decreasing order, and ask for an estimate of the n^{th} encountered value, noted $\Lambda[n]$; in particular $\Lambda[1] = x_{max}$.
- The distribution p_n of $\Lambda[n]$ can be obtained in full generality as:

$$p_n(\Lambda[n]) = N C_{N-1}^{n-1}\, p(x = \Lambda[n]) \left(P(x > \Lambda[n])\right)^{n-1} \left(P(x < \Lambda[n])\right)^{N-n} \longrightarrow (28)$$

- The previous expression means that one has first to choose $\Lambda[n]$ among N variables (N ways), $n-1$ variables among the $N-1$ remaining as the $n-1$ largest ones (C_{N-1}^{n-1} ways), and then assign the corresponding probabilities to the configuration where $n-1$ of them are larger than $\Lambda[n]$ and $N-n$ are smaller than $\Lambda[n]$.
- One can study the position $\Lambda^*[n]$ of the maximum of p_n, and also the width of p_n, defined from the second derivative of $\log p_n$ calculated at $\Lambda^*[n]$.
- The calculation simplifies in the limit where $N \to \infty$, $n \to \infty$, with the ratio n/N fixed.
- In this limit one finds a relation which generalizes equation 18:

$$P_>[\Lambda^*[n]] = n/N \longrightarrow (29)$$

- The width w_ξ of the distribution is found to be given by:

$$w_\xi = \frac{1}{\sqrt{N}} \frac{\sqrt{1-(n/N)^2}}{p(x = \Lambda^*[n])} \longrightarrow (30).$$

Which shows that in the limit $N \to \infty$, the value of the n^{th} variable is more and more sharply peaked around its most probable value given by equation 29.

- In the case of an exponential tail, one finds that $\Lambda^*[n] \simeq \log(N/n)/\alpha$; whereas in the case of power-law tails, one rather obtains:

$$\Lambda^*[n] \simeq A_+ \left(\frac{N}{n}\right)^{1/\xi} \longrightarrow (31)$$

- This last equation shows that, for power-law variables, the encountered values are hierarchically organized: for example, the ratio of the largest value $x_{max} \equiv \Lambda[1]$ to the second largest $\Lambda[2]$ is of the order of $2^{1/\xi}$, which becomes larger and larger as ξ decreases, and conversely tends to 1 when $\xi \longrightarrow \infty$.

- The property in equation 31 is very useful in identifying empirically the nature of the tails of a probability distribution.
- One sorts in decreasing order the set of observed values $\{x_1, x_2, \ldots, x_N\}$ and one simply draws $\Lambda[n]$ as a function of n.
- If the variables are power-law distributed, this graph should be a straight line in log-log plot, with a slope $-1/\xi$ as given by equation 31.
- But such a plot for exponentially distributed variables will show an approximately straight line, but with an effective slope which varies with the total number of points N: the slope is less and less as N/n grows larger.
- In this sense, the formal remark made earlier that an exponential distribution could be seen as a power law with $\xi \longrightarrow \infty$, becomes somewhat more concrete.

- Let us finally note another property of power-laws, potentially interesting for their empirical determination.
- If one computes the average value of x conditioned to a certain minimum value Λ:

$$\langle x \rangle_\Lambda = \frac{\int_\Lambda^\infty x p(x)\, dx}{\int_\Lambda^\infty p(x)\, dx} \longrightarrow (32)$$

then if $p(x)$ decreases as in equation 8, one finds, for $\Lambda \to \infty$

$$\langle x \rangle_\Lambda = \frac{\xi}{\xi - 1} \Lambda \longrightarrow (33)$$

independently of the tail amplitude A_+^{ξ}; This means that ξ can be determined by a one-parameter fit alone.
- The average $\langle x \rangle_\Lambda$ is thus always of the same order as Λ itself, with a proportionality factor which diverges as $\xi \to 1$.

---

* LARGE DEVIATIONS: The Central Limit Theorem teaches us that the Gaussian approximation is justified to describe the "central" part of the distribution of the sum of a large number of random variables (of finite variance).
- However, the definition of the center is rather vague, we haven't said anything quantitative about it.
- The central limit theorem only states that the probability of finding an event in the tails goes to zero for large N.
- Let us now characterize more precisely the region where the Gaussian approximation is valid.

- If X is the sum of N iid random variables of mean m and variance σ^2, one defines a rescaled variable U as:

$$U = \frac{X - Nm}{\sigma\sqrt{N}} \longrightarrow (34)$$

which according to central limit theorem tends towards a Gaussian variable of zero mean and unit variance.
- Hence, for any fixed u, one has:

$$\lim_{N \to \infty} P_{>}(u) = P_{G>}(u) \longrightarrow (35)$$

where $P_{G>}(u)$ is related to the error function and describes the weight contained in the tails of the Gaussian:

$$P_{G>}(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)\, du = \frac{1}{2} \operatorname{erfc}\left(\frac{u}{\sqrt{2}}\right) \longrightarrow (36)$$

- However, the above convergence is not uniform.
- The value of N such that the approximation $P_{>}(u) \simeq P_{G>}(u)$ becomes valid depends on u.
- Conversely, for fixed N, this approximation is only valid for N ~~to~~ not too large: $|u| \ll u_0(N)$.

- One can estimate $u_0(N)$ in the case where the elementary distribution $p_1(x_1)$ is 'narrow', i.e, decreasing faster than any power-law when $|x_1| \to \infty$, such that all moments are finite.
- In this case, all cumulants of p_1 are finite and it becomes possible to obtain a systematic expansion in powers of $N^{-1/2}$ of the difference $\Delta P_{>}(u) \equiv P_{>}(u) - P_{G>}(u)$.

$$\Delta P_>(u) \simeq \frac{\exp(-u^2/2)}{\sqrt{2\pi}}\left(\frac{Q_1(u)}{N^{1/2}} + \frac{Q_2(u)}{N} + \dots + \frac{Q_k(u)}{N^{k/2}}\right) \longrightarrow (37)$$

where the $Q_k(u)$ are polynomial functions which can be expressed in terms of the normalized cumulants of the elementary distribution.

- More explicitly, the first two terms are given by:

$$Q_1(u) = \frac{1}{6}\lambda_3(u^2-1) \longrightarrow (38)$$

and

$$Q_2(u) = \frac{1}{72}\lambda_3^2 u^5 + \frac{1}{8}\left(\frac{1}{3}\lambda_4 - \frac{10}{9}\lambda_3^2\right)u^3 + \left(\frac{5}{24}\lambda_3^2 - \frac{1}{8}\lambda_4\right)u \longrightarrow (39)$$

- One recovers the fact that if all the cumulants of $p_1(x_1)$ of order larger than two are zero, all the Q_k are also identically zero and so is the difference between $p(x,N)$ and the Gaussian.

- For a general symmetric elementary distribution p_1, λ_3 is non-zero.
- The leading term in the above expansion when N is large is thus $Q_1(u)$.
- For the Gaussian approximation to be meaningful, one must at least require that this term is small in the central region where u is of order one, which corresponds to $x - mN \sim \sigma\sqrt{N}$.
- This then imposes that $N \gg N^* = \lambda_3^2$.
- The Gaussian approximation remains valid whenever the relative error is small compared to 1.
- For large u (which will be justified for large N), the relative error is obtained by dividing equation 37 by $P_{G>}(u) \simeq \exp(-u^2/2)/(u\sqrt{2\pi})$.

- One then obtains the following condition:

$$\lambda_3 u^3 \ll N^{1/2} \text{ i.e } |x - Nm| \ll \sigma\sqrt{N}\left(\frac{N}{N^*}\right)^{1/6} \longrightarrow 40$$

- This shows the central region has an extension growing as $N^{2/3}$.

- A symmetric elementary distribution is such that $\lambda_3 \equiv 0$ (because it has no skewness); it is then the kurtosis $\kappa = \lambda_4$ that fixes the first correction to the Gaussian when N is large, and thus the extension of the central region.
- The conditions now read: $N \gg N^* = \lambda_4$ and

$$\lambda_4 u^4 \ll N \text{ i.e } |x - Nm| \ll \sigma\sqrt{N}\left(\frac{N}{N^*}\right)^{1/4} \longrightarrow 41$$

- The central region now extends over a region of width $N^{3/4}$.

- These results don't directly apply if the elementary distribution $p_1(x_1)$ decreases as a power-law (broad distribution).
- In this case, some of the cumulants are infinite and the above cumulant expansion, equation 37, is meaningless.
- This is so because the 'central region' is much more restricted than in the case of narrow distributions.
- We won't explore in any detail, but it is possible to describe the case of 'truncated' power-law distributions, where the above conditions become asymptotically relevant.
- These laws however may have a very large kurtosis, which depends on the point where the truncation becomes noticeable, and the above condition $N \gg \lambda_4$ can be hard to satisfy.

** **CRAMER FUNCTION:** More generally, when N is large, one can write the distribution of the sum of N iid random variables as (we assume their mean is zero, which can always be achieved through a suitable shift of x_i):

$$p(x,N) \underset{N\to\infty}{\simeq} \exp\left[-NS\left(\frac{x}{N}\right)\right] \longrightarrow (42)$$

where S is the so called Cramer's function, which gives some information about the probability of X even outside the 'central' region.

- When the variance is finite, S grows as $S(u) \propto u^2$ for small u's, which again leads to a Gaussian central region.
- For finite u, S can be computed using Laplace's saddle point method, valid for N large.
- By definition:

$$p(x,N) = \frac{1}{2\pi}\int \exp N\left(-iz\frac{x}{N} + \log[\hat{p}_1(z)]\right) dz \longrightarrow (43)$$

- When N is large, the above integral is dominated by the neighborhood of the point z^* where the term in the exponential is stationary.
- The results can be re-written as:

$$p(x,N) \simeq \exp\left[-NS\left(\frac{x}{N}\right)\right] \longrightarrow (44)$$

with $S(u)$ given by:

$$\left.\frac{d\log[\hat{p}_1(z)]}{dz}\right|_{z=z^*} = iu \qquad S(u) = -iz^*u + \log[\hat{p}_1(z^*)] \longrightarrow (45)$$

which, in principle, allows one to estimate $p(x, N^*)$ even outside the central region.

- Note that if $S(u)$ is finite for finite u, the corresponding probability is exponentially small in N.

- In closing, a couple of points:

1) It is my personal opinion that physicists ought to learn Extreme value Statistics as part of their training in Statistical Physics, but this does not seem to be the norm.
- As goes with any opinion, there are no right or wrong answers. I've covered the topic merely based on my viewpoint that has ~~evol~~ evolved from personal experience.

2) The topic of large deviations will certainly come in handy at a later stage when we explore nonlinear extensions to the Fluctuation-Dissipation theorem via the so-called Fluctuation Relations.