Physics 116C
The Distribution of the Sum of Random Variables

Peter Young

(Dated: December 2, 2013)

Consider a random variable with a probability distribution $P(x)$. The distribution is normalized, *i.e.*

$$\int_{-\infty}^{\infty} P(x)\, dx = 1, \tag{1}$$

and the average of x^n (the n-the moment) is given by

$$\boxed{\langle x^n \rangle = \int_{-\infty}^{\infty} x^n P(x)\, dx\,.} \tag{2}$$

The mean, μ, and variance, σ^2, are given in terms of the first two moments by

$$\mu \equiv \langle x \rangle, \tag{3a}$$

$$\sigma^2 \equiv \langle x^2 \rangle - \langle x \rangle^2\,. \tag{3b}$$

The standard deviation, a common measure of the width of the distribution, is just the square root of the variance, *i.e.* σ.

Note: We indicate by angular brackets, $\langle \cdots \rangle$, an average over the exact distribution $P(x)$.

A commonly studied distribution is the Gaussian,

$$P_{\text{Gauss}} = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]\,. \tag{4}$$

We have studied Gaussian integrals before and so you should be able to show that the distribution is normalized, and that the mean and standard deviation are μ and σ respectively. The width of the distribution is σ, since the value of $P(x)$ at $x = \mu \pm \sigma$ has fallen to a fixed fraction, $1/\sqrt{e} \simeq 0.6065$, of its value at $x = \mu$. The probability of getting a value which deviates by n times σ falls off very fast with n as shown in the table below.

n	$P(\lvert x-\mu\rvert > n\sigma)$
1	0.3173
2	0.0455
3	0.0027

In statistics, we often meet problems where we pick N random numbers x_i (this set of numbers we call a "sample") from a distribution and are interested in the statistics of the *mean* $\overline{x}$ of this sample, where

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{5}$$

Note: we use the notation $\overline{\cdots}$ to indicate an average over a sample of data. This is to be compared with $\langle \cdots \rangle$ which indicates an average over the exact distribution. The "law of large numbers", which we will make more precise later, states that the difference between these two averages is small if the sample size N is large.

The x_i might, for example, be data points which we wish to average over. We would be interested in knowing how the sample average $\overline{x}$ deviates from the true average over the distribution $\langle x \rangle$. In other words, we would like to find the distribution of the sample mean $\overline{x}$ if we know the distribution of the individual points $P(x)$. The distribution of the sample mean would tell us about the expected scatter of results for $\overline{x}$ obtained if we repeated the determination of the sample of N numbers x_i many times.

We will actually find it convenient to determine the distribution of the sum

$$\boxed{X = \sum_{i=1}^{N} x_i,} \tag{6}$$

and then then trivially convert it to the distribution of the mean at the end. We will denote the distribution of the sum of N variables by $P_N(X)$ (so $P_1(X) \equiv P(X)$).

We consider the case where the distribution of all the x_i are the same (this restriction can easily be lifted) and that the distribution of the x_i are *statistically independent* (which is not easily lifted). The latter condition means that there are no *correlations* between the numbers, so $P(x_i, x_j) = P(x_i)P(x_j)$ for $i \neq j$.

One can determine the mean and variance of the distribution of the sum in terms of the mean and variance of the individual random variables quite generally as follows. We have

$$\mu_X = \langle \sum_{i=1}^{N} x_i \rangle = N\langle x \rangle = \boxed{N\mu.} \tag{7}$$

$$\sigma_X^2 = \langle \left(\sum_{i=1}^{N} x_i \right)^2 \rangle - \langle \left(\sum_{i=1}^{N} x_i \right) \rangle^2 \tag{8a}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \right) \tag{8b}$$

$$= \sum_{i=1}^{N} \left(\langle x_i^2 \rangle - \langle x_i \rangle^2 \right) \tag{8c}$$

$$= N \left(\langle x^2 \rangle - \langle x \rangle^2 \right) \tag{8d}$$

$$\boxed{= N\sigma^2 \, .} \tag{8e}$$

To get from Eq. (8b) to Eq. (8c) we note that, for $i \neq j$, $\langle x_i x_j \rangle = \langle x_i \rangle \langle x_j \rangle$ since x_i and x_j are assumed to be statistically independent. (This is where the statistical independence of the data is needed.) In other words, quite generally, the mean and variance of the sum of N identically distributed, independent random variables are given by

$$\boxed{\mu_X = N\mu, \qquad \sigma_X^2 = N\,\sigma^2 \, ,} \tag{9}$$

where μ and σ^2 are the mean and variance of the individual variables.

Next we compute the whole distribution of X, which we call $P_N(x)$. This is obtained by integrating over all the x_i with the constraint that $\sum x_i$ is equal to the prescribed value X, *i.e.*

$$\boxed{P_N(X) = \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P(x_1) \cdots P(x_N) \delta(x_1 + x_2 + \cdots x_N - X) \, .} \tag{10}$$

We can't easily do the integrals because of the constraint imposed by the delta function. As discussed in class and in the handout on singular Fourier transforms, we can eliminate this constraint by going to the Fourier transform[1] of $P_N(X)$, which we call $G_N(k)$ and which is defined by

$$\boxed{G_N(k) = \int_{-\infty}^{\infty} e^{ikX} P_N(X) \, dX.} \tag{11}$$

Substituting for $P_N(X)$ from Eq. (10) gives

$$G_N(k) = \int_{-\infty}^{\infty} e^{ikX} \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P(x_1) \cdots P(x_N) \delta(x_1 + x_2 + \cdots x_N - X) \, dX \, . \tag{12}$$

The integral over X is easy and gives

$$G_N(k) = \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_N \, P(x_1) \cdots P(x_N) e^{ik(x_1 + x_2 + \cdots + x_N)}, \tag{13}$$

[1] In statistics the Fourier transform of a distribution is called its "characteristic function".

which is just the product of N identical *independent* Fourier transforms of the single-variable distribution $P(x)$, *i.e.*

$$G_N(k) = \left[\int_{-\infty}^{\infty} P(t)e^{ikt}\,dt\right]^N, \tag{14}$$

or

$$\boxed{G_N(k) = G(k)^N\,,} \tag{15}$$

where $G(k)$ is the Fourier transform of $P(x)$.

Hence to determine the distribution of the sum, $P_N(X)$, the procedure is:

1. Fourier transform the single-variable distribution $P(x)$ to get $G(k)$.

2. Determine $G_N(k)$, the Fourier transform of $P_N(X)$, from $G_N(k) = G(k)^N$.

3. Perform the inverse Fourier transform on $G_N(k)$ to get $P_N(X)$.

Let's see what this gives for the special case of the Gaussian distribution shown in Eq. (4). The Fourier transform $G_{\text{Gauss}}(k)$ is given by

$$\begin{aligned} G_{\text{Gauss}}(k) &= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] e^{ikx}\,dx, \\ &= \frac{e^{ik\mu}}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-t^2/2\sigma^2}\,e^{ikt}\,dt, \\ &= e^{ik\mu}e^{-\sigma^2k^2/2}\,. \end{aligned} \tag{16}$$

To get the second line we made the substitution $x - \mu = t$ and to get the third line we "completed the square" in the exponent, as discussed in class.

$\boxed{\text{The Fourier transform of a Gaussian is therefore a Gaussian.}}$

Equation (15) gives

$$G_{\text{Gauss},N}(k) = G_{\text{Gauss}}(k)^N = e^{iNk\mu}e^{-N\sigma^2k^2/2}, \tag{17}$$

which you will see is the same as $G_{\text{Gauss}}(k)$ except that $\boxed{\mu \text{ is replaced by } N\mu}$ and $\boxed{\sigma^2 \text{ is replaced by } N\sigma^2}$. Hence, when we do the inverse transform to get $P_{\text{Gauss},N}(X)$, we must get a Gaussian as in Eq. (4) apart from these replacements[2], *i.e.*

$$\boxed{P_{\text{Gauss},N}(X) = \frac{1}{\sqrt{2\pi N}\,\sigma}\exp\left[-\frac{(X-N\mu)^2}{2N\sigma^2}\right].} \tag{18}$$

[2] Of course, one can also do the integral explicitly by completing the square to get the same result.

In other words, if the distribution of the individual data points is Gaussian, the distribution of the sum *is also Gaussian* with mean and standard deviation given by Eq. (9).

We now come to an important theorem, the **central limit theorem**, which will be derived in the handout "*Proof of the central limit theorem in statistics*" (using the same methods as above, *i.e.* using Fourier transforms). It applies to to *any* distribution for which the mean and variance exist (*i.e.* are finite), not just a Gaussian. The data must be statistically independent. It states that:

- For any N, the mean of the distribution of the sum is N times the mean of the single-variable distribution (shown by more elementary means in Eq. (7) above).
- For any N, the variance of the distribution of the sum is N times the variance of the single-variable distribution (shown by more elementary means in Eq. (8e) above).
- **(This is the new bit.)** For $N \to \infty$, **the distribution of the sum, $\boldsymbol{P_N(X)}$, becomes Gaussian**, *i.e.* is given by Eq. (18) even if the single-variable distribution $P(x)$ is *not* a Gaussian. This amazing result is is the reason why the Gaussian distribution plays such an important role in statistics.

We will illustrate the convergence to a Gaussian distribution with increasing N for the rectangular distribution

$$P_{\text{rect}}(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & (|x| < \sqrt{3}), \\ 0, & (|x| > \sqrt{3}), \end{cases} \tag{19}$$

where the parameters have been chosen so that $\mu = 0, \sigma = 1$. This is shown by the dotted line in the figure below. Fourier transforming gives

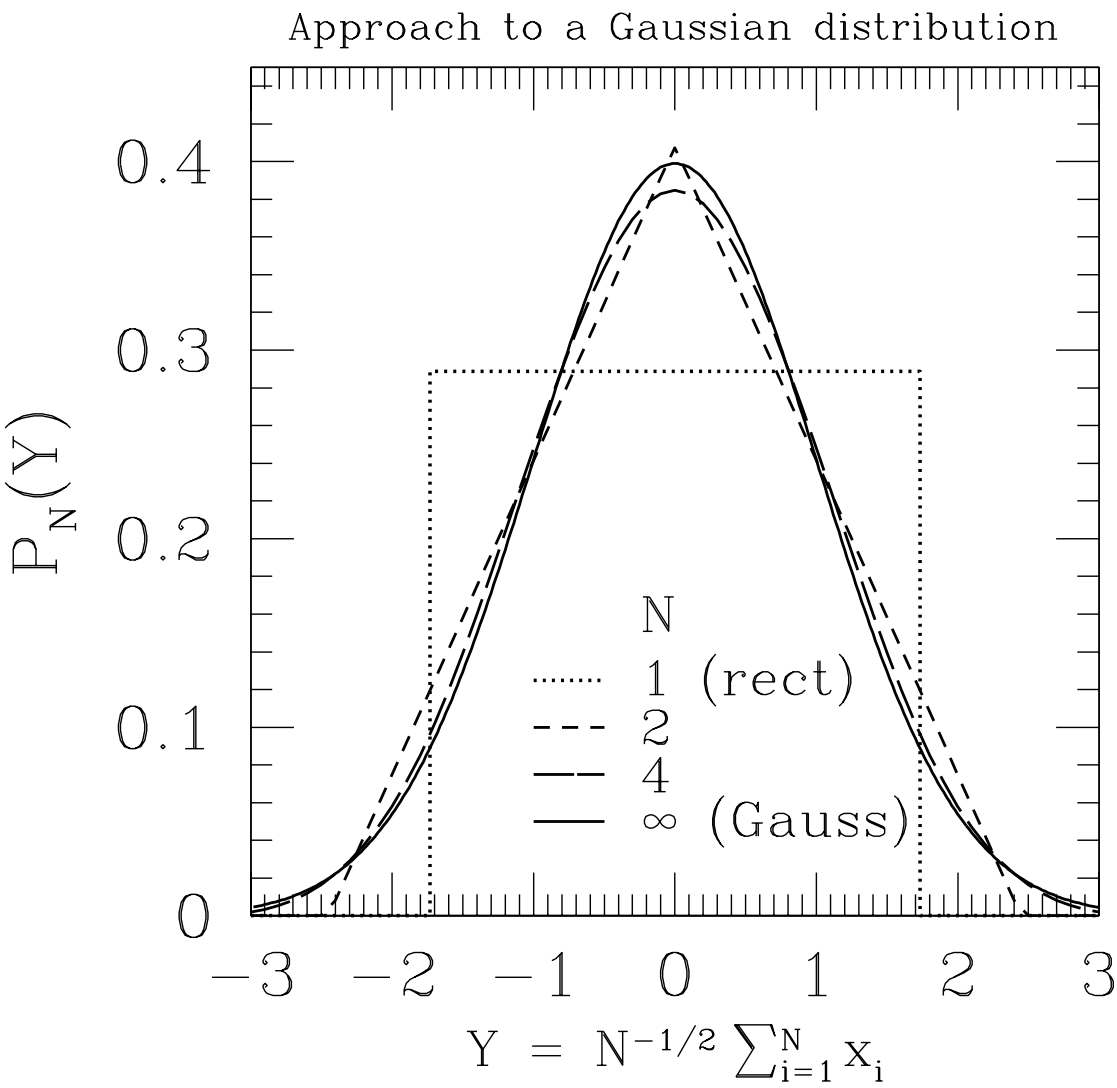
$$\begin{aligned} Q(k) &= \frac{1}{2\sqrt{3}} \int_{-\sqrt{3}}^{\sqrt{3}} e^{ikx}\, dk = \frac{\sin(\sqrt{3}k)}{\sqrt{3}k} \\ &= 1 - \frac{k^2}{2} + \frac{3k^4}{40} - \cdots . \\ &= \exp\left[-\frac{k^2}{2} - \frac{k^4}{20} - \cdots \right] . \end{aligned} \tag{20}$$

Hence

$$P_N(X) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikX} \left(\frac{\sin \sqrt{3}k}{\sqrt{3}k} \right)^N dk,$$

which can be evaluated numerically.

As we have seen, quite generally $P_N(k)$ has mean $N\mu$, ($= 0$ here), and standard deviation $\sqrt{N}\sigma$ ($= \sqrt{N}$ here). Hence, to illustrate that the distributions really do tend to a Gaussian for $N \to \infty$ I plot below the distribution of $Y = X/\sqrt{N}$ which has mean 0 and standard deviation 1 for all N.



Results are shown for $N = 1$ (the original distribution), $N = 2$ and 4, and the Gaussian ($N = \infty$). The approach to a Gaussian for large N is clearly seen. Even for $N = 2$, the distribution is much close to Gaussian than the original rectangular distribution, and for $N = 4$ the difference from a Gaussian is quite small on the scale of the figure. This figure therefore provides a good illustration of the central limit theorem.

Equation (18) can also be written as a distribution for the sample mean $\overline{x}$ $(= X/N)$ as[3]

$$\boxed{P_{\text{Gauss},N}(\overline{x}) = \sqrt{\frac{N}{2\pi}}\,\frac{1}{\sigma}\exp\left[-\frac{N(\overline{x}-\mu)^2}{2\sigma^2}\right].} \tag{22}$$

Let us denote the mean (obtained from many repetitions of choosing the set of N numbers x_i) of the sample average by $\mu_{\overline{x}}$, and the standard deviation of the sample average distribution by $\sigma_{\overline{x}}$. Equation (22) tells us that

$$\boxed{\mu_{\overline{x}} = \mu,} \tag{23a}$$

$$\boxed{\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{N}}.} \tag{23b}$$

Hence the mean of the distribution of sample means (averaged over many repetitions of the data) is the exact mean μ, and its standard deviation, which is a measure of its width, is $\sigma/\sqrt{N}$ which becomes small for large N.

These statements tell us that an average over many independent measurements will be close to the exact average, intuitively what one expects[4]. For example, if one tosses a coin, it should come up heads on average half the time. However, if one tosses a coin a small number of times, N, one would not expect to necessarily get heads for *exactly* $N/2$ of the tosses. Six heads out of 10 tosses (a fraction of 0.6), would intuitively be quite reasonable, and we would have no reason to suspect a biased toss. However, if one tosses a coin a million times, intuitively the same fraction, 600,000 out of 1,000,000 tosses, would be most unlikely. From Eq. (23b) we see that these intuitions are correct because the characteristic deviation of the sample average from the true average (1/2 in this case) goes down proportional to $1/\sqrt{N}$.

To be precise, we assign $x_i = 1$ to heads and $x_i = 0$ to tails so $\langle x_i\rangle = 1/2$ and $\langle x_i^2\rangle = 1/2$ and hence, from Eqs. (3),

$$\mu = \frac{1}{2}, \qquad \sigma = \sqrt{\frac{1}{2} - \left(\frac{1}{2}\right)^2} = \sqrt{\frac{1}{4}} = \frac{1}{2}. \tag{24}$$

---

[3] We obtained Eq. (22) from Eq. (18) by the replacement $X = N\overline{x}$. In addition the factor of $1/\sqrt{N}$ multiplying the exponential in Eq. (18) becomes $\sqrt{N}$ in Eq. (22). Why did we do this? The reason is as follows. If we have a distribution of y, $P_y(y)$, and we write y as a function of x, we want to know the distribution of x which we call $P_x(x)$. Now the probability of getting a result between x and $x+dx$ must equal the probability of a result between y and $y+dy$, *i.e.* $P_y(y)|dy| = P_x(x)|dx|$. In other words

$$P_x(x) = P_y(y)\left|\frac{dy}{dx}\right|. \tag{21}$$

As a result, when transforming the distribution of X in Eq. (18) into the distribution of $\overline{x}$ one needs to multiply by N. You should verify that the factor of $|dy/dx|$ in Eq. (21) preserves the normalization of the distribution, *i.e.* $\int_{-\infty}^{\infty} P_y(y)dy = 1$ if $\int_{-\infty}^{\infty} P_x(x)dx = 1$.

[4] Intuitive statements of this type are loosely called "the law of large numbers".

For a sample of N tosses, Eqs. (23) gives the sample mean and standard deviation to be

$$\mu_{\overline{x}} = \frac{1}{2}, \qquad \sigma_{\overline{x}} = \frac{1}{2\sqrt{N}}\,. \tag{25}$$

Hence the magnitude of a typical deviation of the sample mean from 1/2 is $\sigma_{\overline{x}} = 1/(2\sqrt{N})$. For $N = 10$ this is about 0.16, so a deviation of 0.1 is quite possible (as discussed above), while for $N = 10^6$ this is 0.0005, so a deviation of 0.1 (200 times $\sigma_{\overline{x}}$) is most unlikely (also discussed above). In fact we can calculate the probability of a deviation (of either sign) of 200 or more times $\sigma_{\overline{x}}$ since the central limit theorem tells us that the distribution is Gaussian for large N. As discussed in class, the result can be expressed as a complementary error function, i.e.

$$\frac{2}{\sqrt{2\pi}} \int_{200}^{\infty} e^{-t^2/2}\, dt = \operatorname{erfc}\left(\frac{200}{\sqrt{2}}\right) = 5.14 \times 10^{-2629}, \tag{26}$$

(where erfc is the complementary error function), *i.e. very* unlikely!