Physics 116C

Proof of the central limit theorem in statistics.

Peter Young

(Dated: November 23, 2009)

In this handout we give a proof of the central limit theorem, which we have already discussed.

Consider a random variable with a probability distribution $P(x)$. The mean, μ, and variance, σ^2, are given by

$$\begin{aligned}\mu &\equiv \langle x\rangle = \int_{-\infty}^{\infty} x\,P(x)\,dx \\ \sigma^2 &\equiv \langle x^2\rangle - \langle x\rangle^2 \,.\end{aligned}$$

The standard deviation is just the square root of the variance, *i.e.* σ. In this handout we consider distributions that fall off sufficiently fast at ∞ that the mean and variance are finite. This *excludes*, for example, the Lorentzian distribution:

$$P_{\rm Lor} = \frac{1}{\pi}\frac{1}{1+x^2}\,. \tag{1}$$

A common distribution which *does* have a finite mean and variance is the Gaussian distribution

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

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.

Consider a distribution, *not necessarily Gaussian*, with a finite mean and distribution. An example would be the rectangular distribution

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

where the parameters have been chosen so that $\mu = 0, \sigma = 1$. We pick N random numbers x_i from such a distribution and form the sum

$$X = \sum_{i=1}^{N} x_i .$$

We are interested to determine the distribution of X, which we call $P_N(X)$. For example, if $N = 2$, we know that if the sum is X then x_2 must equal $X - x_1$. Hence the distribution of X is the product $P(x_1)P(X - x_1)$ integrated over x_1, *i.e.*

$$P_2(X) = \int_{-\infty}^{\infty} P(x_1)P(X - x_1)\, dx_1 \,. \tag{4}$$

You will recognize this as a convolution and recall that the Fourier transform of a convolution is the product of the Fourier transforms of the individual functions. Hence, if $Q(k)$ is the Fourier transform of $P(x)$ (in the context of statistics the Fourier transform of a distribution is called its *characteristic function*), and $Q_N(k)$ is the the Fourier transform of $P_N(X)$, we have

$$Q_2(k) = Q(k)^2 \,, \tag{5}$$

where

$$Q(k) = \int_{-\infty}^{\infty} P(x)e^{ikx}\, dx \,, \qquad Q_2(k) = \int_{-\infty}^{\infty} P_2(X)e^{ikx}\, dx \,.$$

Expanding out the exponential we can write $Q(k)$ in terms of the moments of $P(x)$

$$Q(k) = 1 + ik\langle x\rangle + \frac{(ik)^2}{2!}\langle x^2\rangle + \frac{(ik)^3}{3!}\langle x^3\rangle + \cdots \,.$$

It will be convenient in what follows to write $Q(k)$ as an exponential, *i.e.*

$$\begin{aligned} Q(k) &= \exp\left[\ln\left(1 + ik\langle x\rangle + \frac{(ik)^2}{2!}\langle x^2\rangle + \frac{(ik)^3}{3!}\langle x^3\rangle + \cdots\right)\right] \\ &= \boxed{\exp\left[ik\mu - \frac{k^2\sigma^2}{2} + c_3(ik)^3 + c_4(ik)^4 + \cdots\right] \,,} \end{aligned} \tag{6}$$

where c_3 involves third and lower moments, c_4 involves fourth and lower moments, and so on.

We illustrate this for the special case of a Gaussian. As discussed before, the Fourier transform of a Gaussian is also a Gaussian, a result which was obtained by "completing the square". Completing the square on Eq. (2) gives

$$\boxed{Q_{\text{Gauss}}(k) = \exp\left[ik\mu - \frac{k^2\sigma^2}{2}\right] \,,} \tag{7}$$

showing that the higher order coefficients, c_3, c_4, etc. in Eq. (6) *all vanish* for a Gaussian.

Consider now the case $N > 2$. If the sum of the x_i is X we can choose the first $N - 1$ as we wish but then x_N must equal $X - \sum_{i=1}^{N-1} x_i$. Hence Eq. (4) is generalized to

$$P_N(X) = \int_{-\infty}^{\infty} P(x_1)dx_1 \int_{-\infty}^{\infty} P(x_2)dx_2 \cdots \int_{-\infty}^{\infty} P(x_{N-1})dx_{N-1}\, P[X - (x_1 + x_2 + \cdots + x_{N-1})] \,.$$

This is a also known as a convolution, generalized to the case of N variables. It can easily be shown, using the methods discussed in class which led to Eq. (5), that the the Fourier transform of this generalized convolution is again the product of the Fourier transforms of the individual functions. This means that

$$Q_N(k) = Q(k)^N = \exp\left[ikN\mu - \frac{Nk^2\sigma^2}{2} + Nc_3(ik)^3 + Nc_4(ik)^4 + \cdots\right].$$

Comparing with Eq. (6) we see that

> the mean of the distribution of the sum of N variables (the coefficient of $-ik$ in the exponential is N times the mean of the distribution of one variable, and the variance of the distribution of the sum of N variables (the coefficient of $-k^2/2$) is N times the variance of the distribution of one variable.

These are general statements applicable for *any* N.

However, if N is *large* a great simplification can be made. The distribution of X, which is the inverse transform of $Q_N(k)$, is given by

$$\begin{aligned} P_N(X) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} Q(k)^N\, e^{-ikX}\, dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left[-ikX' - \frac{Nk^2\sigma^2}{2} + Nc_3(ik)^3 + Nc_4(ik)^4 + \cdots\right] dk\,, \end{aligned} \tag{8}$$

where

$$X' = X - N\mu\,. \tag{9}$$

Looking at the $-Nk^2/2$ term in the exponential in Eq. (8), we see that the integrand is significant for $k < k^\star$, where $N\sigma^2(k^\star)^2 = 1$, and negligibly small for $k \gg k^\star$. However, for $0 < k < k^\star$ the higher order terms in Eq. (8), (*i.e.* those of order k^3, k^4 etc.) are very small since $N(k^\star)^3 \sim N^{-1/2}, N(k^\star)^4 \sim N^{-1}$ and so on. Hence the terms of higher order than k^2 in Eq. (8), do not contribute for large N and so

$$\lim_{N\to\infty} P_N(X) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left[-ikX' - \frac{Nk^2\sigma^2}{2}\right] dk\,. \tag{10}$$

In other words, for large N the distribution is the Fourier transform of a Gaussian, which, as we know, is also a Gaussian. Completing the square in Eq. (10) gives

$$\begin{aligned} \lim_{N\to\infty} P_N(X) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left[-\frac{N\sigma^2}{2}\left(k - \frac{iX'}{N\sigma^2}\right)^2\right] dk\ \exp\left[-\frac{(X')^2}{2N\sigma^2}\right] \\ &= \boxed{\frac{1}{\sqrt{2\pi N}\,\sigma}\exp\left[-\frac{(X-N\mu)^2}{2N\sigma^2}\right]}\,, \end{aligned} \tag{11}$$

where, in the last line, we used Eq. (9). This is a Gaussian with mean $N\mu$ and variance $N\sigma^2$. Eq. (11) is the central limit theorem in statistics. It tells us that,

> for $N \to \infty$, the distribution of the sum of N variables is a Gaussian of mean N times the mean, μ, of the distribution of one variable, and variance N times the variance of the distribution of one variable, σ^2, independent of the form of the distribution of one variable, $P(x)$, provided only that μ and σ are finite.

The central limit theorem is of such generality that it is extremely important. It is the reason why the Gaussian distribution has such a preeminent place in the theory of statistics.