Orthogonality Does Not Imply Asymptotic Normality — DA Freedman
Statistics 215 — October 2007

Consider random variables which are orthogonal, with mean 0 and variance 1, and uniformly bounded fourth moments. The CLT need not hold—i.e., the sum need not be asymptotically normal—because independence is not assumed. (The CLT, being a theorem, remains true.) We present several examples, normalizing the sum $S_n = X_1 + \cdots X_n$ by $\sqrt{n}$, or by

$$D_n = \sqrt{\sum_{j=1}^{n} X_j^2}.$$

The examples are relevant to general forms of the OLS model $Y = M\beta + \epsilon$ that require only $\mathrm{cov}(\epsilon|M) = \sigma^2 I_{n\times n}$ rather than IID errors; normalization by D_n is akin to normalizing regression statistics by $\hat{\sigma}$. The punchline: the usual asymptotics need not hold for the OLS estimator $\hat{\beta}$, without the assumption of IID errors given M. (We write M for the design matrix to avoid confusion with the random variables X_j.)

$(*)$ Conditions. The X_j have $E(X_j) = 0$ and $E(X_j^2) = 1$. Furthermore $E(X_j^4)$ is uniformly bounded, and $E(X_j X_k) = 0$ for $j \neq k$.

Example 1. Let Z and $\{Y_j\}$ be independent. Suppose the Y_j are independent, $E(Y_j) = 0$, $E(Z^2) = E(Y_j^2) = 1$. Finally, suppose $E(Z^4)$ and $E(Y_j^4)$ are finite. Let $X_j = ZY_j$. These random variables satisfy the conditions $(*)$, but $S_n/\sqrt{n}$ isn't asymptotically normal: the limiting distribution is normal, multiplied by Z. Normalizing by D_n does give asymptotic normality, because Z cancels.

Example 2. Let $U_j = 0$ or $\sqrt{2}$ be a sequence of random variables constructed as follows. With probability 1/2, the sequence consists of a long block of 0's, followed by a very long block of $\sqrt{2}$'s, followed by a very very long block of 0's, etc. With the remaining probability 1/2, the 0's and $\sqrt{2}$'s are interchanged. The Y_j are IID ± 1 with probability 1/2, independent of $\{U_j\}$. Let $X_j = U_j Y_j$. Again, these random variables satisfy the conditions $(*)$. Clearly, $\max_j U_j^2 = 2$ and $D_n^2 \to \infty$, so $\max_{j\le n} U_j^2 = o(D_n^2)$. Furthermore, $\mathrm{var}(S_n|U) = D_n^2$. Thus, S_n/D_n is asymptotically normal. With rapidly increasing block length, D_n^2/n oscillates between $0+$ and $2-$. So $S_n/\sqrt{n}$ isn't asymptotically normal.

Our next example involves $\xi_j = \sin(j\theta)$, where θ is uniform on the circle $[0, 2\pi)$ and j is an integer; the main interest is $j = 1, 2, \ldots$. If $z_j = \exp(ij\theta)$ with $i = \sqrt{-1}$ and $\exp(z) = e^z$, then ξ_j is the imaginary part of $z_j = \cos(j\theta) + i\sin(j\theta)$. The next lemma follows by computing moments.

Lemma 2. The z_j all have the same distribution; furthermore, for each j, as $n \to \infty$, the joint distribution of z_n and z_j converges weak-star, the two variables becoming independent.

Lemma 3.

(i) $E(\xi_j) = 0$; in fact, all odd moments vanish.

(ii) $E(\xi_j^2) = 1/2$ and $E(\xi_j^4) = 3/8$.

(iii) $E(\xi_j \xi_k) = 0$ for $j \neq k$.

(iv) $\sum_{j=1}^{n} \cos(j\theta)$ is the real part of

$$\Psi_n(\theta) = \frac{e^{(n+1)i\theta} - e^{i\theta}}{e^{i\theta} - 1},$$

and $\sum_{j=1}^{n} \sin(j\theta)$ is the imaginary part.

(v) $\sum_{j=1}^{n} \sin^2(j\theta) = \frac{1}{2}(n - q_n)$ where $q_n = \sum_{j=1}^{n} \cos(2j\theta)$ is the real part of $\Psi_n(2\theta)$.

Example 3. Let $X_j = \sqrt{2}\xi_j$. Conditions $(*)$ are satisfied. However, S_n converges in distribution, and D_n^2 is of order n. Whether we normalize S_n by $\sqrt{n}$ or D_n—or not at all—there is no asymptotic normality.

Sourav Chatterjee suggested that examples could be based on U-statistics. For $\ell = 1, 2, \ldots,$ let the U_ℓ be IID, with $P(U_\ell = \pm 1) = 1/2$. Let

$$Q_n = \sum_{1 \leq j \neq k \leq n} U_j U_k = \Big(\sum_{\ell=1}^{n} U_\ell\Big)^2 - \Big(\sum_{\ell=1}^{n} U_\ell^2\Big)$$

whose distribution is asymptotic to $n(\chi_1^2 - 1)$. Note that $Q_{n+1} - Q_n = 2\big(\sum_{\ell=1}^{n} U_\ell\big)U_{n+1}$.

Example 4. Let $T_j = \sum_{\ell=1}^{j} U_\ell / \sqrt{j}$ and let $X_{j+1} = T_j U_{j+1}$ for $j = 1, 2, \ldots$. Let $X_1 = U_1$. Conditions $(*)$ are easily verified; for the rest, we rely on simulations. To begin with, S_n is very skewed to the right, so cannot be asymptotically normal. On the other hand, S_n/D_n—although not far from normal—has a negative mean. We can replace T_j by $f(T_j)$ for suitable functions f, although $\mathrm{var}\, f(T_j)$ may then depend a little on j and n. If $f(x) = x^6$, then S_n itself has a much longer tail than the normal; indeed, S_n is roughly like a symmetrized log normal variable. By contrast, S_n/D_n is short-tailed and bimodal. (Numerator and denominator are somewhat dependent.) Neither $S_n/\sqrt{n}$ nor S_n/D_n is asymptotically normal.

Back-of-the-envelope arguments suggest

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} f\Big(\frac{1}{\sqrt{j}} \sum_{\ell=1}^{j} U_\ell\Big) U_{j+1} \to \int_0^1 f(B_t/\sqrt{t})\, dB_t \tag{1}$$

$$\frac{1}{n} D_n^2 = \frac{1}{n} \sum_{j=1}^{n} f\Big(\frac{1}{\sqrt{j}} \sum_{\ell=1}^{j} U_\ell\Big)^2 \to \int_0^1 f(B_t/\sqrt{t})^2\, dt \tag{2}$$

where B is Brownian motion. Because the summands are uncorrelated,

$$\operatorname{var}\left\{\frac{1}{\sqrt{n}}\sum_{j=1}^{n/K} f\Big(\frac{1}{\sqrt{j}}\sum_{\ell=1}^{j} U_\ell\Big)U_{j+1}\right\} = \frac{1}{n}\sum_{j=1}^{n/K}\operatorname{var}\left\{f\Big(\frac{1}{\sqrt{j}}\sum_{\ell=1}^{j} U_\ell\Big)\right\} = O\left(\frac{1}{K}\right) \tag{3}$$

Likewise,

$$E\left\{\frac{1}{n}\sum_{j=1}^{n/K}\Big[f\Big(\frac{1}{\sqrt{j}}\sum_{\ell=1}^{j} U_\ell\Big)\Big]^2\right\} = \frac{1}{n}\sum_{j=1}^{n/K} E\left\{\Big[f\Big(\frac{1}{\sqrt{j}}\sum_{\ell=1}^{j} U_\ell\Big)\Big]^2\right\} = O\left(\frac{1}{K}\right) \tag{4}$$

The singularity near 0 therefore seems unimportant.

Example 5. (Chaterjee.) Let the summands be $U_j U_k$, with $j < k$ ordered by k and within k by j. Summands are identically distributed, taking the values ± 1 with probability 1/2 each. The summands have mean 0; the square of each summand is identically 1; summands are orthogonal. Consider the subsequence $m_n = n(n+1)/2$ where the sum S_{m_n} is

$$\sum_{1\le j<k\le n} U_j U_k = \frac{1}{2}\Big[\Big(\sum_{\ell=1}^{n} U_\ell\Big)^2 - \Big(\sum_{\ell=1}^{n} U_\ell^2\Big)\Big] \sim \frac{1}{2}n(\chi_1^2 - 1)$$

We normalize by $\sqrt{n(n+1)/2} \doteq n/\sqrt{2}$; the limiting distribution is $(\chi_1^2 - 1)/\sqrt{2}$.

Convergence seems to hold along the full sequence of n's. More specifically, suppose $m_n \le m \le m_{n+1}$. The normalizing $\sqrt{m} \sim n$. The difference between the sum at m and at $n(n+1)/2$ is a sum of order n terms, which is $O_P(\sqrt{n}) = O_P(\sqrt[4]{m})$. After division by $\sqrt{m}$, the difference is $O_P(1/\sqrt[4]{m})$.

The distribution of $(\chi_1^2 - 1)/\sqrt{2}$ has mean 0 and variance 1, but is longer-tailed than N(0,1). For instance,

$$P\big\{(\chi_1^2 - 1)/\sqrt{2} > 2.6\big\} = P\big\{Z^2 > 1 + 2.6\sqrt{2}\big\} = .03,$$

while $P\big\{|Z| > 2.6\big\} = .009$, where Z is N(0,1). The tail area is off by a factor of 3, and it gets worse further out. On the other hand, annoyingly,

$$P\big\{(\chi_1^2 - 1)/\sqrt{2} > 2\big\} = P\Big\{|Z| > \sqrt{1 + 2\sqrt{2}}\Big\} \doteq P\{|Z| > 1.96\} \doteq .05,$$

The first probability is one-sided:

$$P\{(\chi_1^2 - 1)/\sqrt{2} < -2\} = P\{Z^2 < 1 - 2\sqrt{2}\} = 0,$$

but the symmetric tail area is very close.

Steve Evans has another construction, which gives a sequence $X_1, X_2, \ldots$ of uncorrelated random variables having mean 0 and variance 1, with subsequences of

$$\mathcal{L}\big\{(X_1 + \cdots + X_n)/\sqrt{n}\big\}$$

close to any distribution with mean 0 and variance 1.

For regression asymptotics assuming independent errors, see

http://www.stat.berkeley.edu/users/census/Ftest.pdf