A SPATIAL EXAMPLE OF THE MULTIVARIATE CENTRAL LIMIT THEOREM

Consider a set of only two adjacent spatial locations, $\{s_1, s_2\}$, and suppose that a set of unobserved *spatially depend site attributes*, $e_i = (e_{1i}, e_{2i})'$, $i = 1,...,n$, are each generated by dependent sums of dichotomous ("fair coin-flip") events:

(1) $$e_{1i} = x_{1i} + x_{2i}$$

(2) $$e_{2i} = x_{3i} + x_{2i}$$

where (x_{1i}, x_{2i}, x_{3i}), $i = 1,..,n$ are all independently and identically distributed with $\Pr(x_{ki} = 1) = 1/2$, $k = 1,2,3$, $i = 1,..,n$. [Here we could use values (-1,1) to ensure that $E(e_{1i}) = E(e_{2i}) = 0$, but we have chosen to leave them as simple sums of "coin flips".] With these definitions,

(3) $$E(e_{1i}) = E(x_{1i}) + E(x_{2i}) = .5 + .5 = 1$$

(4) $$\mathrm{var}(e_{1i}) = \mathrm{var}(x_{1i}) + \mathrm{var}(x_{2i}) = .5(.5) + .5(.5) = .50$$

and similarly for e_{2i}. Moreover, the covariance between e_{1i} and e_{2i} is of the form,

(5) $$\mathrm{cov}(e_{1i}, e_{2i}) = E(e_{1i}e_{2i}) - E(e_{1i})E(e_{2i}) = E(e_{1i}e_{2i}) - (1)(1)$$

where

(6) $$E(e_{1i}e_{2i}) = E[(x_{1i} + x_{2i})(x_{3i} + x_{2i})] = E(x_{1i}x_{3i}) + E(x_{2i}x_{3i}) + E(x_{1i}x_{2i}) + E(x_{2i}^2)$$

$$= (.5)^2 + (.5)^2 + (.5)^2 + [\mathrm{var}(x_{2i}) + E(x_{2i})^2] = .75 + [.25 + (.5)^2 = 1.25$$

Hence this covariance is given by

(7) $$\mathrm{cov}(e_{1i}, e_{2i}) = .25 > 0$$

which in turn quantifies the degree of spatial dependency in terms of the *correlation*

(8) $$\rho(e_{1i}, e_{2i}) = \frac{\mathrm{cov}(e_{1i}, e_{2i})}{\sigma(e_{1i})\sigma(e_{2i})} = \frac{.25}{\sqrt{.5}\sqrt{.5}} = \frac{.25}{.5} = .5$$

Then if the unobserved residuals at locations $\{s_1, s_2\}$ are the cumulative sums of these unobserved attributes, say

$$(9) \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} e_{1i} \\ e_{2i} \end{pmatrix} = \sum_{i=1}^{n} e_i$$

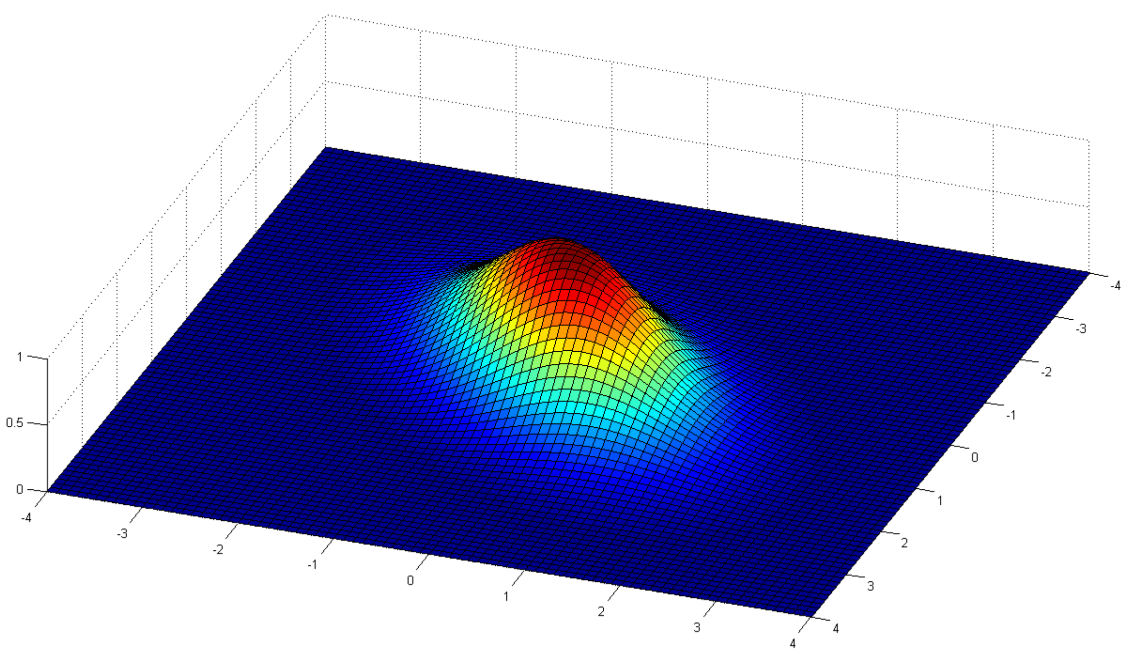
then the *Multivariate Central Limit Theorem* says that for large n, the standardized version

$$(10) \qquad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad \text{with} \quad z_i = \frac{\varepsilon_i - E(\varepsilon_i)}{\sigma(\varepsilon_i)} \ , \ i = 1,2$$

of this random vector should be approximately Normally Distributed as $\varepsilon \sim N(0,\Sigma)$ where

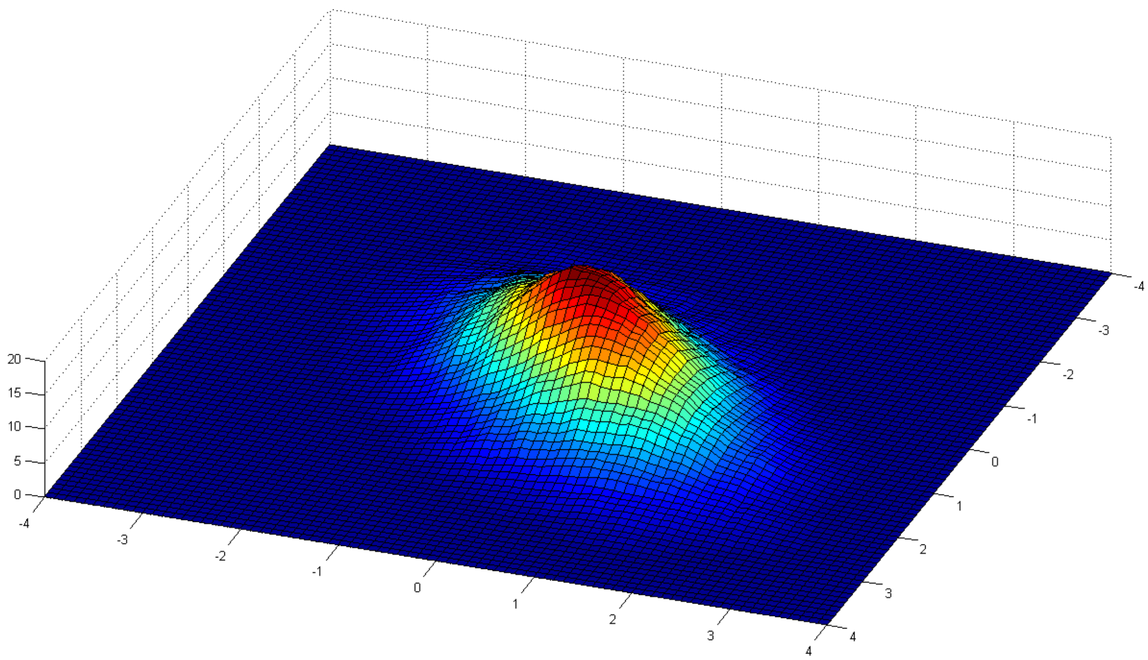
$$(11) \qquad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$$

To illustrate this, a Matlab program, **bi_normal.m**, was written to plot this limit distribution [with the command: **bi_normal (.5)**] as shown below.



Theoretical Limiting Bivariate Normal Distribution

To compare this limit with an actual simulated distribution, a second program, **bi_normal_sim.m**, was written to approximate the sampling distribution of z for various choices of n for a large number of simulated samples (*sims*) from this distribution. For the choice, $n = 100$ and sims = 10,000, the command **bi_normal_sim(100,10000)** yielded the following example:



Simulated Sum of Spatially Dependent Effects

Hence it should be clear from a comparison of these two results that the Multivariate Central Limit Theorem is indeed at work.