

Visualizing Type II Error in Normality Tests

José A. Sánchez-Espigares, Pere Grima, Lluís Marco-Almagro

*Department of Statistics and Operational Research, Universitat Politècnica de Catalunya
-BarcelonaTech, Barcelona, 08028, Spain*

Author's footnote:

José A. Sánchez-Espigares is an Associate Professor at Universitat Politècnica de Catalunya – BarcelonaTech, Barcelona, Spain (e-mail: josep.a.sanchez@upc.edu), Pere Grima is a Professor at Universitat Politècnica de Catalunya – BarcelonaTech, Barcelona, Spain (e-mail: pere.grima@upc.edu); and Lluís Marco-Almagro is an Associate Professor at Universitat Politècnica de Catalunya – BarcelonaTech, Barcelona, Spain (e-mail: lluis.marco@upc.edu)

Visualizing Type II Error in Normality Tests

A Skewed Exponential Power Distribution, with parameters defining kurtosis and skewness, is introduced as a way to visualize Type II error in normality tests. By varying these parameters a mosaic of distributions is built, ranging from double exponential to uniform or from positive to negative exponential; the normal distribution is a particular case located in the center of the mosaic. Using a sequential color scheme, a different color is assigned to each distribution in the mosaic depending on the probability of committing a Type II error. This graph gives a visual representation of the power of the performed test. This way of representing results facilitates the comparison of the power of various tests and the influence of sample size. A script to perform this graphical representation, programmed in the R statistical software, is available online as supplementary material.

Key Words: Normality test, Type II error, Visualizing information, Teaching statistics, SEPD Distributions

1. INTRODUCTION

Generally in hypothesis testing, what we want to verify has the burden of proof. Nevertheless, this is not the case in normality tests, where what we want to prove (the distribution is normal) is assumed as true. Moreover, not having an alternative distribution, the discussion on Type II error (considering the distribution is normal when it is not) is frequently skipped or at least minimized in general statistics courses, although this Type II error can be high.

This problem has been widely discussed in the statistical literature, and several studies on the power of existing tests have been published. For example, Farrell and Rogers-Stewart (2006), Yacini and Yolacan (2007) and Yap and Sim (2011) compare the power

of different normality tests against a set of alternative distributions. Results are shown in tables or power curves based on sample size, and a discussion of the most appropriate test depending on the kind of deviation from normality for each of the alternative distributions is included.

Other papers emphasize the interest of having a visual representation of the test, facilitating the observation of departures from normality. This happens with the Kolmogorov-Smirnov test (KS), which is easy to understand and visualize, and still remains common in both textbooks and software packages (although its power is worse than that of other normality tests). In this regard, Rosenkrantz (2000) derives bounds for the theoretical quantile function suggesting a test that has the same visual representation advantages of the KS test. In a similar manner, Aldor-Noiman et al. (2013) propose a new method which provides simultaneous confidence bands for a normal quantile-quantile plot that are narrower in the extremes than those associated with the KS test, so higher powers are obtained, keeping the visualization benefit.

This paper presents a visual and easy to understand way to show Type II error in normality tests. A graph shows a set of alternative distributions in a grid – a kind of mosaic of distributions – with the normal distribution in the center, and changing asymmetry as we move left and right, and changing kurtosis as we move up and down. A wide range of alternatives is then visible, and it is possible to test the hypothesis of normality for samples generated from each distribution. A sequential color scheme is then used to represent the probability of declaring normality, giving a visual representation of the power of the normality test used.

The following sections develop in detail the construction of the graph, together with an example showing its practical application.

2. MOSAIC OF DISTRIBUTIONS

Zhu and Zinde-Walsh (2009) show (as a particular case of a more general distribution) a version of the Skewed Exponential Power Distribution (SEPD) that can be adapted to our needs:

$$f(x | \mu^*, \sigma^*, \alpha, p) = \begin{cases} \frac{1}{\sigma^*} K(p) \exp\left(-\frac{1}{p} \left|\frac{x - \mu^*}{2\alpha\sigma^*}\right|^p\right) & \text{if } x \leq \mu^* \\ \frac{1}{\sigma^*} K(p) \exp\left(-\frac{1}{p} \left|\frac{x - \mu^*}{2(1-\alpha)\sigma^*}\right|^p\right) & \text{if } x > \mu^* \end{cases} \quad (1)$$

We say that $X \sim \text{SEPD}(\mu^*, \sigma^*, \alpha, p)$, where μ^* and σ^* are respectively the location and scale parameters, and correspond to the mean and standard deviation in the case of the Normal distribution, $p \geq 0$ is the parameter related to kurtosis, $\alpha \in [0, 1]$ is related to skewness and $K(p)$ is the normalization constant, $K(p) = 1/[2p^{1/p}\Gamma(1 + 1/p)]$. Values $p = 2$ and $\alpha = 0.5$ correspond to the Normal distribution.

The kurtosis changes when moving p and keeping all other parameters constant. Analogously, the skewness changes when varying α (Figure 1).

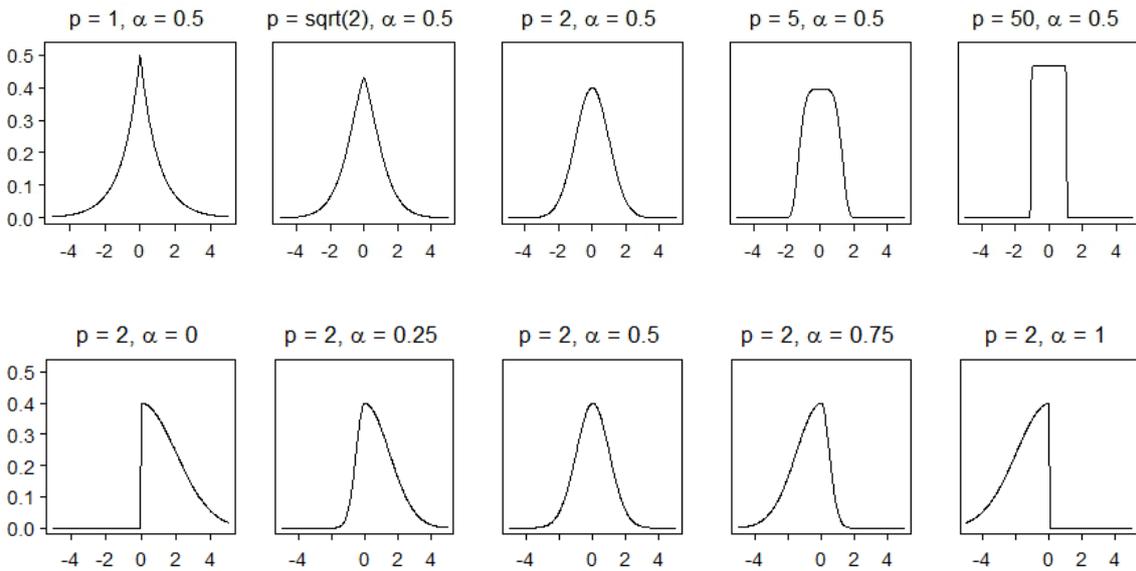


Figure 1: pdf of SEPD with $\mu^* = 0$ and $\sigma^* = 1$, with different values of p keeping $\alpha = 0.5$ (upper row) and different values of α keeping $p = 2$ (lower row)

As we want all distributions to have the same values of μ and σ , consider $X_z \sim SEPD(0; 1; \alpha; p)$. From the expressions obtained by Zhu and Zinde-Walsh (2009) for $E(X)$ and $V(X)$ follows that:

$$E(X_z) = \frac{1}{K(p)} \left[(1 - 2\alpha) \frac{p\Gamma(2/p)}{\Gamma^2(1/p)} \right] \quad (2)$$

$$V(X_z) = \frac{1}{[K(p)]^2} \left[(3\alpha^2 - 3\alpha + 1) \frac{p^2\Gamma(3/p)}{\Gamma^3(1/p)} \right] - \frac{1}{[K(p)]^2} \left[(1 - 2\alpha) \frac{p\Gamma(2/p)}{\Gamma^2(1/p)} \right]^2 \quad (3)$$

We write $E(X_z) = A$ and $V(X_z) = B^2$ to lighten the notation and define:

$$Y \equiv g(X_z) = \mu + \sigma \left(\frac{X_z - A}{B} \right) \quad (4)$$

One can check that $E(Y) = \mu$ and $V(Y) = \sigma^2$. In order to define the probability density function (pdf) of Y , a change of variable is applied:

$$f_Y(y) = f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right|$$

With:

$$f_X(x_z|0, 1, p, \alpha) = \begin{cases} K(p) \exp\left(-\frac{1}{p} \left| \frac{x_z}{2\alpha} \right|^p\right) & \text{if } x \leq 0 \\ K(p) \exp\left(-\frac{1}{p} \left| \frac{x_z}{2(1-\alpha)} \right|^p\right) & \text{if } x > 0 \end{cases}$$

Isolating X_z in (4) we obtain:

$$g^{-1}(y) = \frac{y - \mu}{\sigma} B + A$$

Therefore, we have:

$$f(y) = \begin{cases} \frac{B}{\sigma} K(p) \exp\left(-\frac{1}{p} \left| \frac{A + B \left(\frac{y - \mu}{\sigma}\right)}{2\alpha} \right|^p\right) & \text{if } y \leq \mu - \frac{A}{B} \sigma \\ \frac{B}{\sigma} K(p) \exp\left(-\frac{1}{p} \left| \frac{A + B \left(\frac{y - \mu}{\sigma}\right)}{2(1 - \alpha)} \right|^p\right) & \text{if } y > \mu - \frac{A}{B} \sigma \end{cases} \quad (5)$$

In order to draw the mosaic of distributions, values of p between 1 (double exponential) and 50 (almost an uniform distribution) are used. The normal distribution ($p = 2$) is placed in the center. Each value of p is equal to the previous one raised to a certain value j , except for the second, which will always be equal to $2^{1/j^{(m/2-1.5)}}$, where m is the number of distributions on each side of the mosaic. For instance, for a 7x7 mosaic, these values will be: 1, $2^{1/j^2}$, $2^{1/j}$, 2, 2^j , 2^{j^2} , 2^{j^3} . As the highest value of p is equal to 50, in this case of $m = 7$ it can be deduced that $j = \sqrt[3]{\frac{\log 50}{\log 2}}$. In order to have the Normal distribution in the center, m must be an odd number; therefore, in general, $j = \frac{m}{2} - 0.5 \sqrt{\frac{\log 50}{\log 2}}$.

For α we take equidistant values between 0 and 1 (inclusive). Therefore, if we have n values, the i -th position will be $\alpha = \frac{i-1}{n-1}$

With these criteria it is possible to obtain mosaics of distributions that vary in a reasonably equidistant way between the negative ($\alpha = 0$) and positive ($\alpha = 1$) exponential distributions, and from the double exponential ($p = 1$) to an almost uniform distribution ($p = 50$). Figures 2 and 3 show 7x7 and 25x25 mosaics made with the statistical software R.

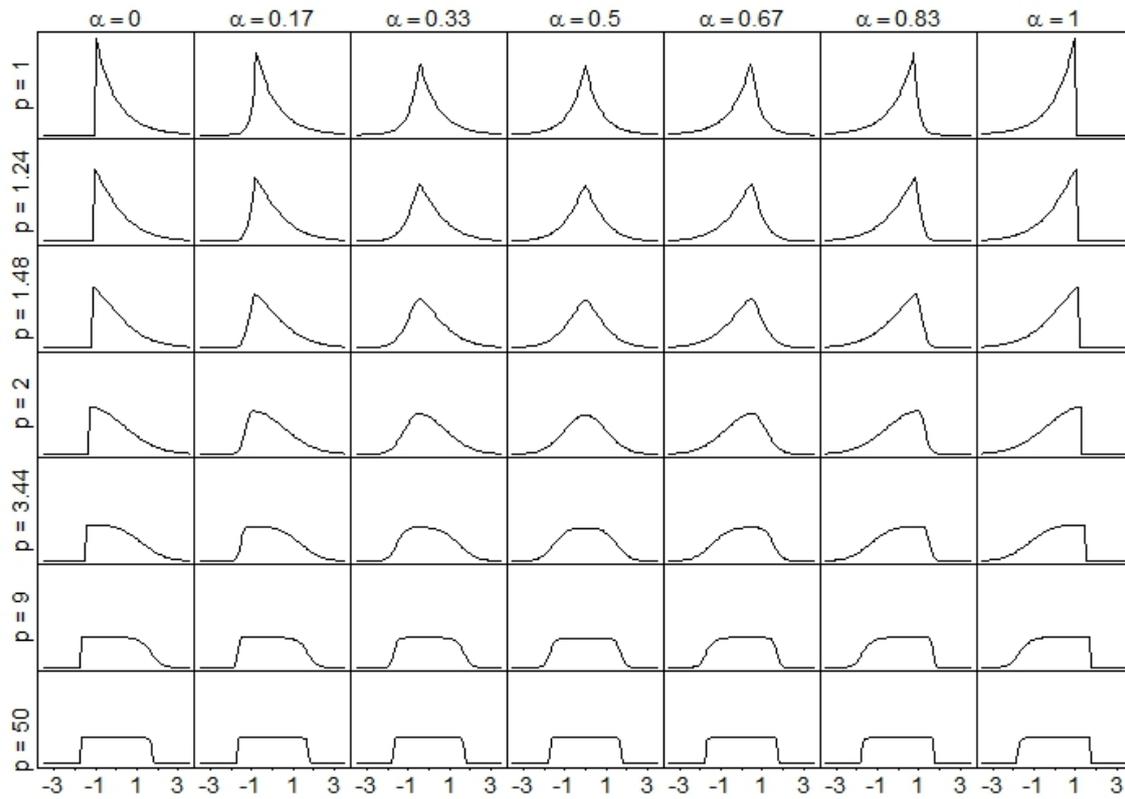


Figure 2: Mosaic of distributions (7x7) with $\mu = 0$, $\sigma = 1$

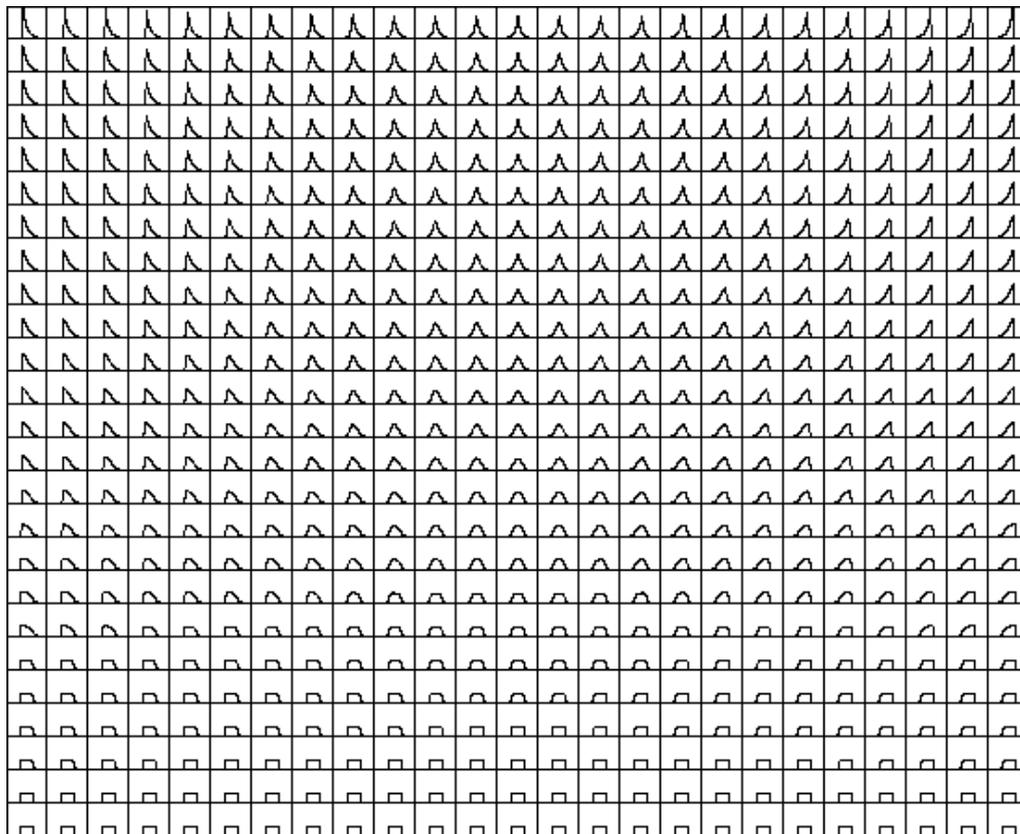


Figure 3: Mosaic of distributions (25x25)

3. VISUALIZING THE PROBABILITY OF DECLARING NORMALITY

Given a sample of size n , N samples of each of the distributions in the mosaic are generated. The values of μ and σ do not affect the appearance of the mosaic as the distributions do not have a scale in the axis. The random number generation for each distribution is done by a procedure analogous to that used by Zhu and Zinde-Walsh (2009). Values following $Y \sim SEPD(0; 1; \alpha; p)$ are generated from a random number U from a uniform distribution $U(0, 1)$, and another random number W from a gamma distribution with shape parameter $k = 1/p$ and scale parameter $\theta = 1$, in the form:

$$Y = \begin{cases} \frac{-\alpha W^{1/p}}{\Gamma\left(1 + \frac{1}{p}\right)} & \text{if } U < \alpha \\ \frac{(1 - \alpha)W^{1/p}}{\Gamma\left(1 + \frac{1}{p}\right)} & \text{if } U \geq \alpha \end{cases}$$

A normality test is performed for each of the N samples of each distribution, and each p -value is recorded. Given a certain level of significance, we assign a value to each distribution equal to the proportion of samples (out of N) for which the assumption of normality is not rejected.

Figure 4 shows 7x7 mosaics with the probability (for each distribution) of declaring normality when using the Shapiro-Wilk test. These values are based on simulation of $N = 10000$ samples for each distribution. A level of significance of 5% is used. The calculations were performed with the R statistical software, using the 'nortest' package (Gross, 2013). Obviously, the power of the test increases with the sample size n .

A sequential gray scale is used to display results more clearly. From lighter to darker colors the ranges are 0-0.05, 0.05-0.25, 0.25-0.50 and 0.50-1.

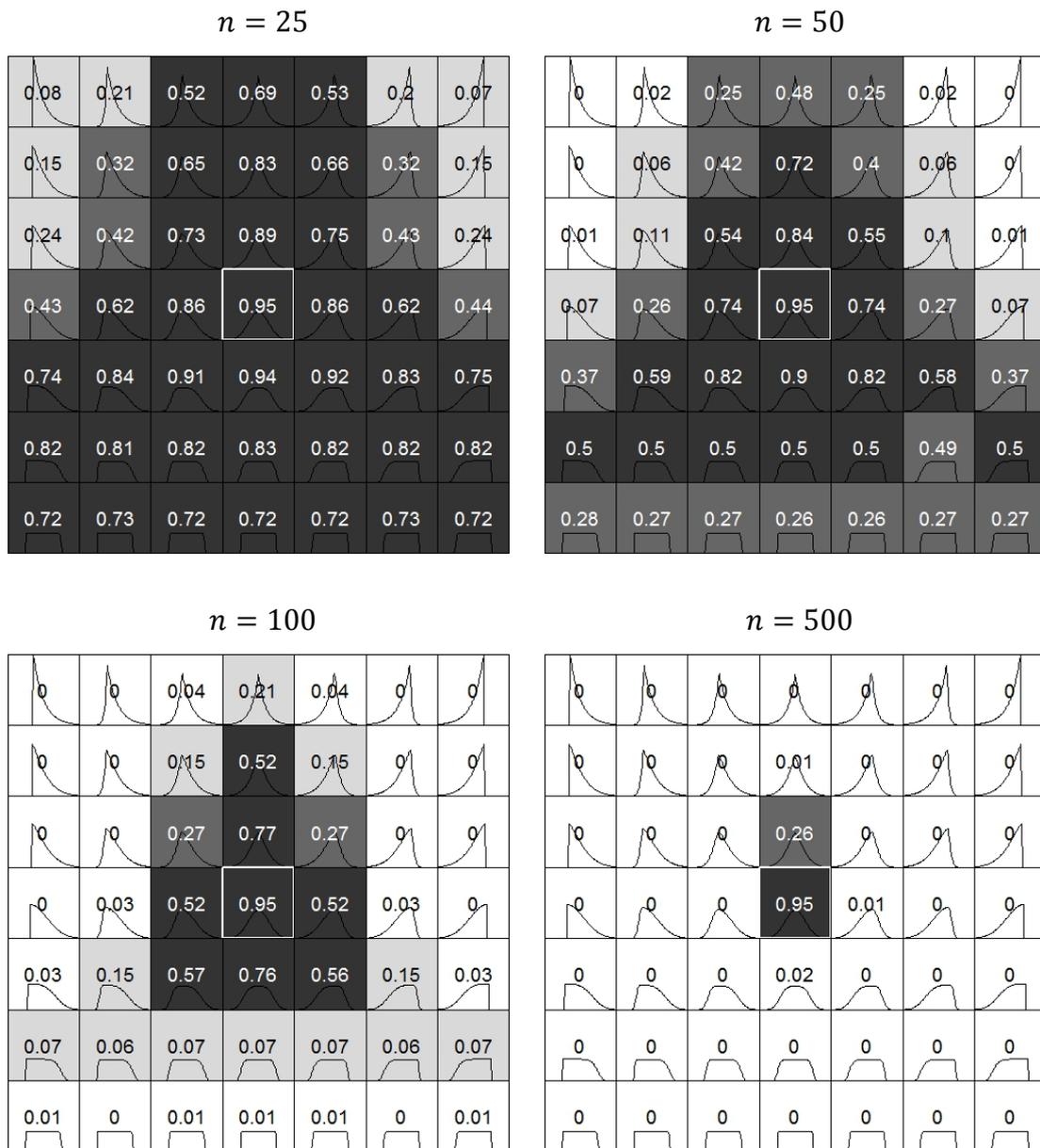


Figure 4: Mosaic of distributions including the proportion of times that the hypothesis of normality is not rejected (with 0.05 as significance level) using the Shapiro-Wilk test for the sample sizes shown

Figure 5 shows the results obtained using the following normality tests: Kolmogorov-Smirnov-Lilliefords, Anderson-Darling and Shapiro-Wilk, with samples of size $n = 25, 50$ and 100. Dark areas in the mosaics are basically symmetric about the vertical axis crossing the center of the mosaic, due to the fact that distribution shapes are also symmetric about this axis. At the lower lines of the mosaic the distribution is almost uniform, so it changes scarcely from left to right. Therefore, the test either declares or not normality for all distributions in that line. On the contrary, skewness is relevant in

the upper lines of the mosaic, with very asymmetric distributions on the left and the right parts of the graph, so the test only declares normality, at most, in the central distributions. These behaviors give the candle flare shape that can be seen in the graphs.

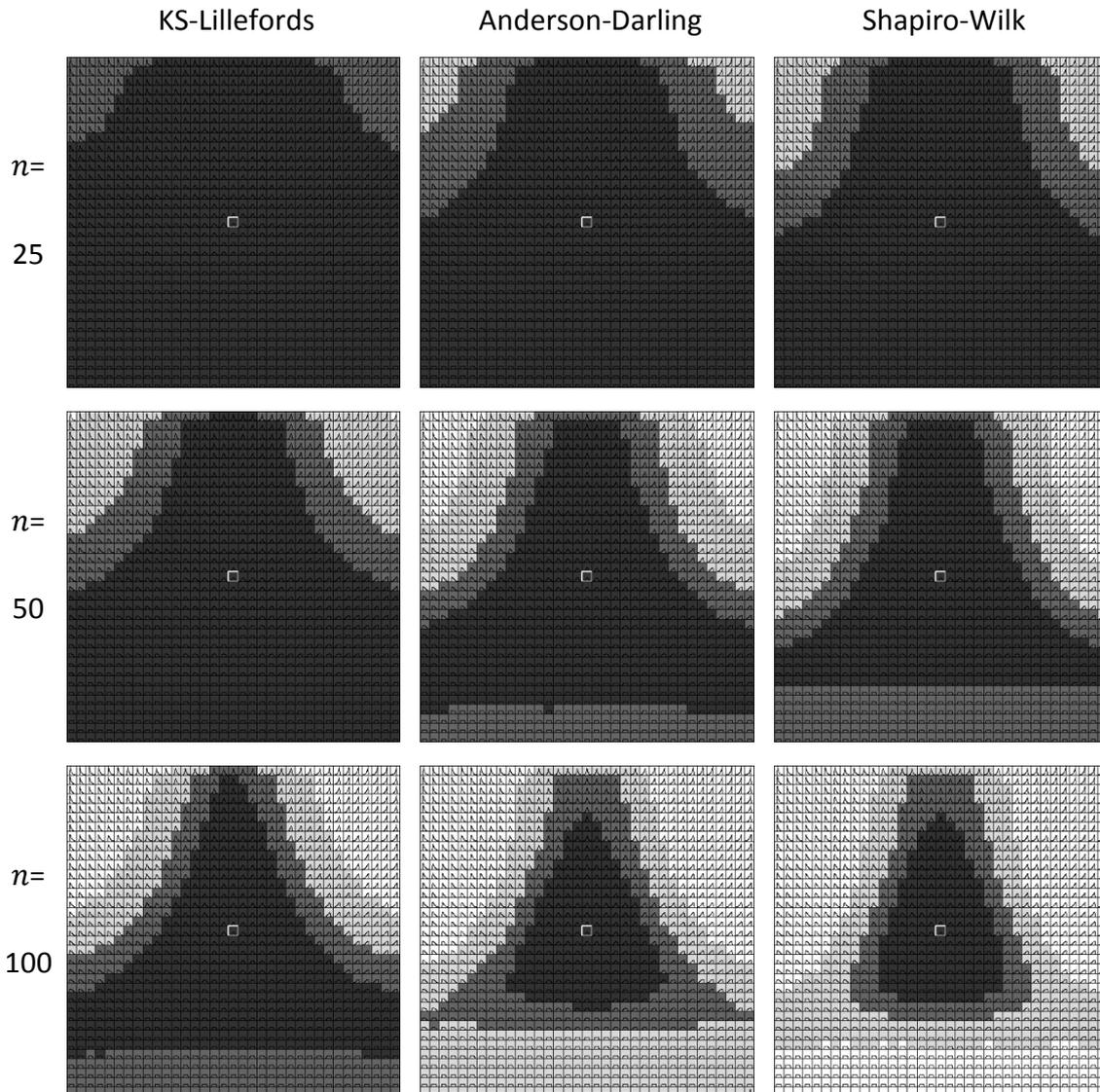


Figure 5: 35x35 mosaic of distributions for comparing the results using three different normality tests with three sample sizes.

The smaller the dark area in the graph, the more powerful the test is. Obviously, when the sample size is increased, the dark area becomes smaller. Looking at Figure 5 we can also conclude that Anderson-Darling test works better than KS-Lilleforde test, and Shapiro-Wilk test better than Anderson-Darling. The ideal situation –unattainable when

working with samples – would be having only dark black in the center, where the normal distribution is located.

In our experience, the existence and importance of Type II error in normality tests is often neglected both in academic and business settings. The graphs shown in this paper allow the visualization of Type II error and, in this way, help in understanding the limitations of these kind of tests. Also, the mosaic facilitates the comparison among normality tests and the impact of sample size in a visual manner.

SUPPLEMENTARY MATERIALS

An R script named “Visualizing Type II Error.R” is attached as supplementary material. The script allows graphical representations as those shown in this paper. Instructions on how to use the script are contained in the script code.

REFERENCES

- Aldor-Noiman, S., Brown, L.D., Buja, A., Rolke, W. and Stine, R.A. (2013), “The Power to See: A New Graphical Test of Normality,” *The American Statistician*, 67, 249-260.
- Farrell, P.J., Rogers-Stewart, K. (2006), “Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test,” *Journal of Statistical Computation and Simulation*, 76, 803-816.
- Gross, J. and bug fixes by Ligges, U. (2012). nortest: Tests for Normality. R package version 1.0-2. <http://CRAN.R-project.org/package=nortest>
- Rosenkrantz, W.A. (2000), “Confidence Bands for Quantile Functions: A Parametric and Graphic Alternative for Testing Goodness of Fit,” *The American Statistician*, 54, 185-190.

Yacini, B., Yolacan, S. (2007), "A comparison of various tests of normality," *Journal of Statistical Computation and Simulation*, 77, 175-183.

Yap, B.W., Sim, C.H. (2011), "Comparisons of various types of normality tests," *Journal of Statistical Computation and Simulation*, 81, 2141-2155.

Zhu, D. and Zinde-Walsh, V. (2009), "Properties and estimation of asymmetric exponential power distribution," *Journal of Econometrics*, 148, 89-99