VALIDATION OF MULTIVARIATE MONTE CARLO STUDIES.

Jorge Luis Romeu

Department of Mathematics
SUNY-Cortland
Cortland, NY 13045

Abstract. Validating a simulation study is a complex but necessary process. All study results depend on the strength of the validation statement. In Monte Carlo simulations, validation opportunities become particularly reduced. The multidimensionality issue only increases the problem complexity. In this paper, a three-phase validation scheme based on the multivariate generation methods adopted in the study, is presented and explained in detail. Examples of the implementation of such a scheme, in three large Monte Carlo power studies, are described in detail.

1.0 Introduction

Recent computing advances (e.g., evergrowing power of PC's, parallel processing) have spurred the use of Monte Carlo techniques in statistical work. From engineering applications (Romeu, 1985), to comparison of methods (Romeu, 1989), to teaching (Romeu, 1986) or methodological research (Romeu, 1988), Monte Carlo and system simulation methods have become an important working tool for the modern statistician.

No longer is the practitioner constrained by the dimension of the problems. Hence, we are increasingly seeing the development of Monte Carlo techniques in the areas of multivariate statistics. And comparisons of multivariate methods (Ozturk and Romeu,

Key words and phrases. Monte Carlo, simulation, multivariate analysis, multivariate generation, statistical distributions..

J. L. Romeu is an Associate Professor of Statistics and Computer Science at SUNY-Cortland and an Adjunct Profeossor of Statistics and Operations Research at Syracuse University (SU). He is a Research Fellow of the CASE Center of SU., where this research was undertaken, under a Dr. Nuala McGann SUNY/UUP Award and a Supercomputer grant from the Cornell Theory Center. Romeu is a Fellow of the Institute of Statisticians and a Member of ASA and ORSA.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

1992), small sample studies (Romeu, 1992a) or validation of statistical theories (Romeu, 1992b) are proliferating by day.

However, Monte Carlo results are only as good as the faith we can have in the validity of such studies. And building this faith becomes increasingly difficult when dealing with multivariate Monte Carlo. For, in addition to the significant differences with system simulation that inhibit the use of specific types of validation techniques, we add the multidimensionality problem.

In this paper, we describe a **three-phase validation scheme** for multivariate Monte Carlo studies. This methodology is derived from our experiences in planning and implementing extensive studies of this type. For example, in Romeu (1990), we compared ten multivariate normality (MVN) Goodness of Fit (GOF) tests (Table 1), under twelve non-normal alternatives (Figure 1 and Table 2). In the comparison we used experimental settings with two, four and eight p-variates, four sample sizes and two covariance structures (Figure 2), for a total of 288 experimental treatments or simulation runs. We will use this and other similar experiences to illustrate the application of this validation methodology.

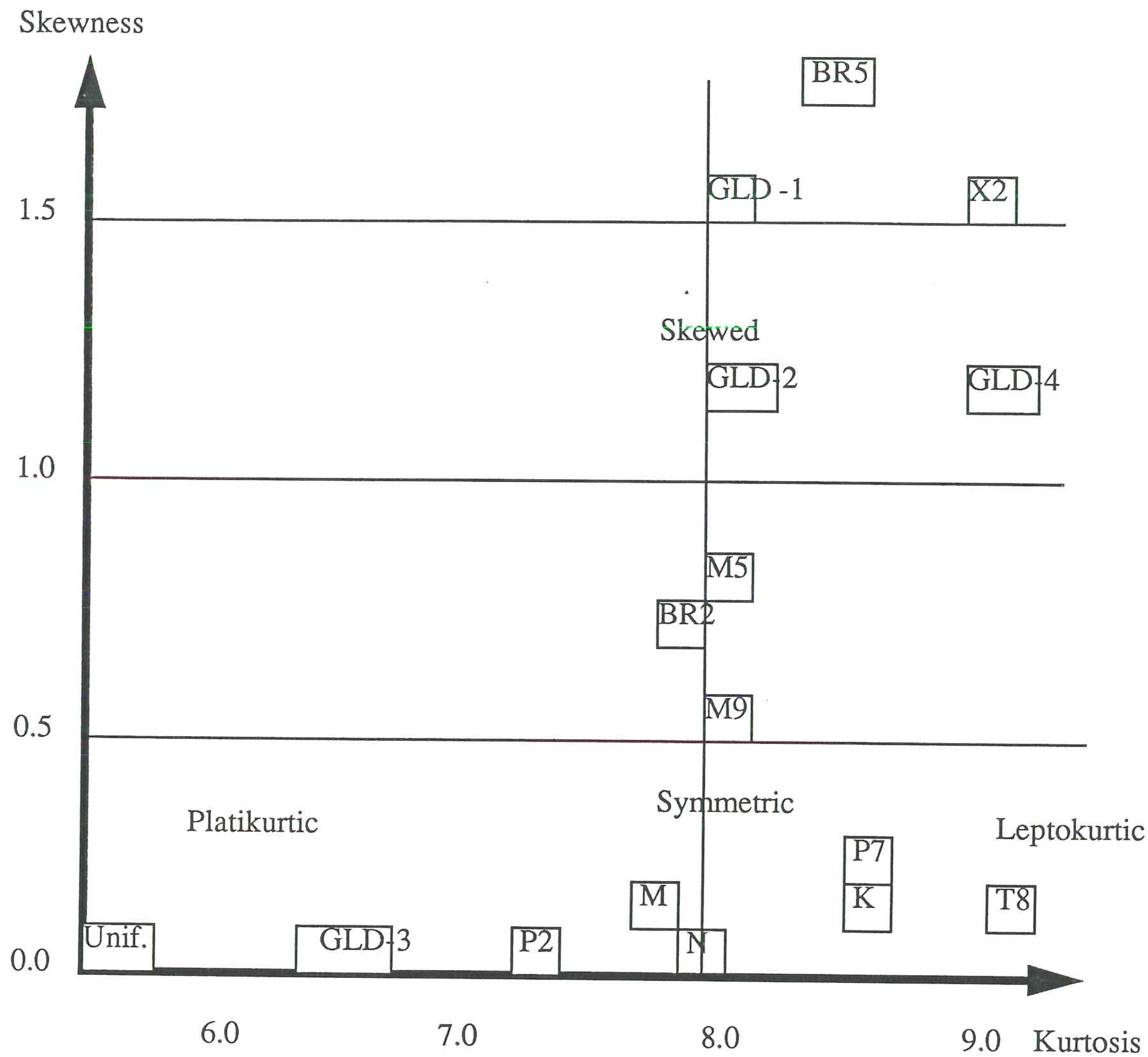
The three proposed validation phases are: (i) *planning*, during the design stage of the study, (ii) *concurrent*, as we move along the study itself and (iii) *final* stage, using the study results. Carried out in such a way, validation becomes a researcher's quality control tool instead of just an activity performed to satisfy a client or a journal reviewer.

For, a good validation methodology prevents that, at the end of hundreds of runs, we find out that our experiment somehow went wrong. And that we could have detected and corrected the problem earlier, if a carefull monitoring (validation) scheme had been implemented.

2.0 Planning Stage

Monte Carlo studies are driven by statistical problems with intractable or messy mathematical solutions. Otherwise, the use of the Monte Carlo approach would be ill-advised. However, there is frequently an associated problem (the asymptotic version, a special case) with a well known closed form solution. It is during the initial literature search, while researching the theory behind the problem, that its associated solved version can be brought out to light. We may also find, during this initial research, that previous work exists in this general area with some reliable numerical results. And these activities will provide

SCHEMATIC OF THE STATISTICAL ALTERNATIVES DESIGN, WITH RESPECT TO THEIR RELATION TO SKEWNESS/KURTOSIS. ALTERNATIVES CONSIDERED ARE: Normal (N); Uniform (Unif); t-dist(T8); Chi Square (X2); Morgenstern (M); Kinchine (K); Pearson Type II (P2) and Type VII (P7); Mixtures of Normals (M5, M9) and BivReg(BR).



Normal Distribution has Skewness 0.0 and Kurtosis 8.0

BIVARIATE DISTRIBUTIONS.

FIGURE 1.

Table 1. Multivariate Normality Tests Compared:

==============================

Multivariate Q_n: Cholesky version.

Multivariate Q_n: Sigma Inverse version.

Mardia's Skewness Test.

Mardia's Kurtosis Test.

Cox and Small Test.

Koziol's Angles Test.

Koziol's Chi Square Test.

Malkovich and Afifi's Test.

Royston's Test.

Hawkins' Test.

Table 2. Multivariate Statistical Alternatives:

============================

Bivariate Morgenstern (with two parameters).

Bivariate Kinchine (with two parameters).

Bivariate Regression (with two parameters).

Pearson Type II (with $m = 10, 6, 4$).

Pearson Type VII (with $m = 10, 6, 2$).

Mixtures of Normals (with two mixing parameters).

Student t (with 8 degrees of freedom).

Chi Square (with 10 degrees of freedom).

Generalized Lambda Distribution (three versions).

Uniform (0,1).

SCHEMATIC OF THE MONTE CARLO STUDY EXPERIMENTAL DESIGN TREE:

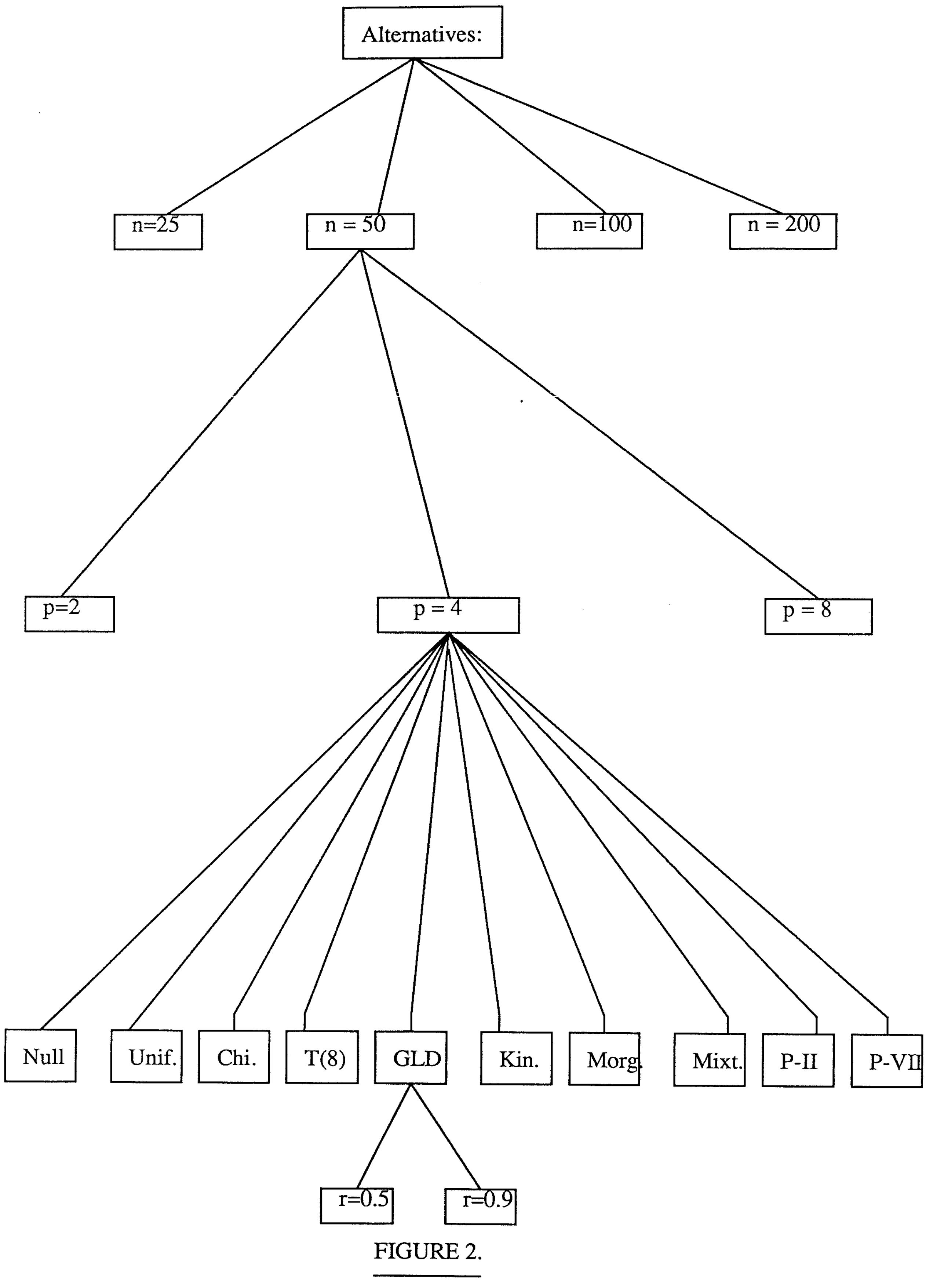


FIGURE 2.

our first validation parameters.

For example, in Romeu (1992a) we studied and compared the small sample properties of ten MVN GOF tests, through our empirically derived small sample critical values. However, asymptotic distributions existed for some of these tests. And we used them to validate our work, by showing how the empirical critical values actually tended to the asymptotic ones, as $n \to \infty$.

We also found, during our literature search, that Mardia (1970, 1979) and Koziol (1982, 1983 and 1986) had obtained limited subsets of empirical critical values for their tests. We used these numerical results to check and validate our work in progress.

Power studies require the generation of well specified types of statistical alternatives. This activity constitutues the main challenge in a multivariate Monte Carlo study. But it also provides one of its most useful validation tools. For, by carefull investigaton of the statistical alternatives used, their properties and their generation methods, we can find additional validation parameters with which to check our work.

In Romeu (1990) we classified the twelve statistical alternatives used in the power study into *purely skewed*, *purely kurtic* and *combined*, based on their first four moments. We also discovered that most MVN GOF tests investigated were either *skewed-prone* or *kurtic-prone*. And we classified them as such. For example, we verified how, in the bivariate skewness vs. kurtosis plane, Pearson Type II distribution yielded zero skewness and kurtosis smaller than that of the bivariate Normal (Figure 2). We realized we could use a combination of Mardia's Skewness and Kurtosis tests, applied to the generated samples, to construct another bidimensional plot (Figure 3). And that we could use these plots as validation tools. For both types of plots graphed the alternative distributions in different ways, but in compatible Skewness vs. Kurtosis planes. Generation methods were validated by verifying that both bivariate distribution classifications (the first plane representing the theoretical and the second the empirical conditions) would be consistent.

Up to now, we have been discussing our use of methods for generating multivariate distributions in the validation procedure. However, multivariate generation is not a trivial problem. And there exist several approaches to it. We surveyed them and organized the material into two broad groups which we call (*i*) indirect approaches to generating multivariate distributions and (*ii*) direct methods.

The indirect approaches are based on combining natural or empirical univariate dis-

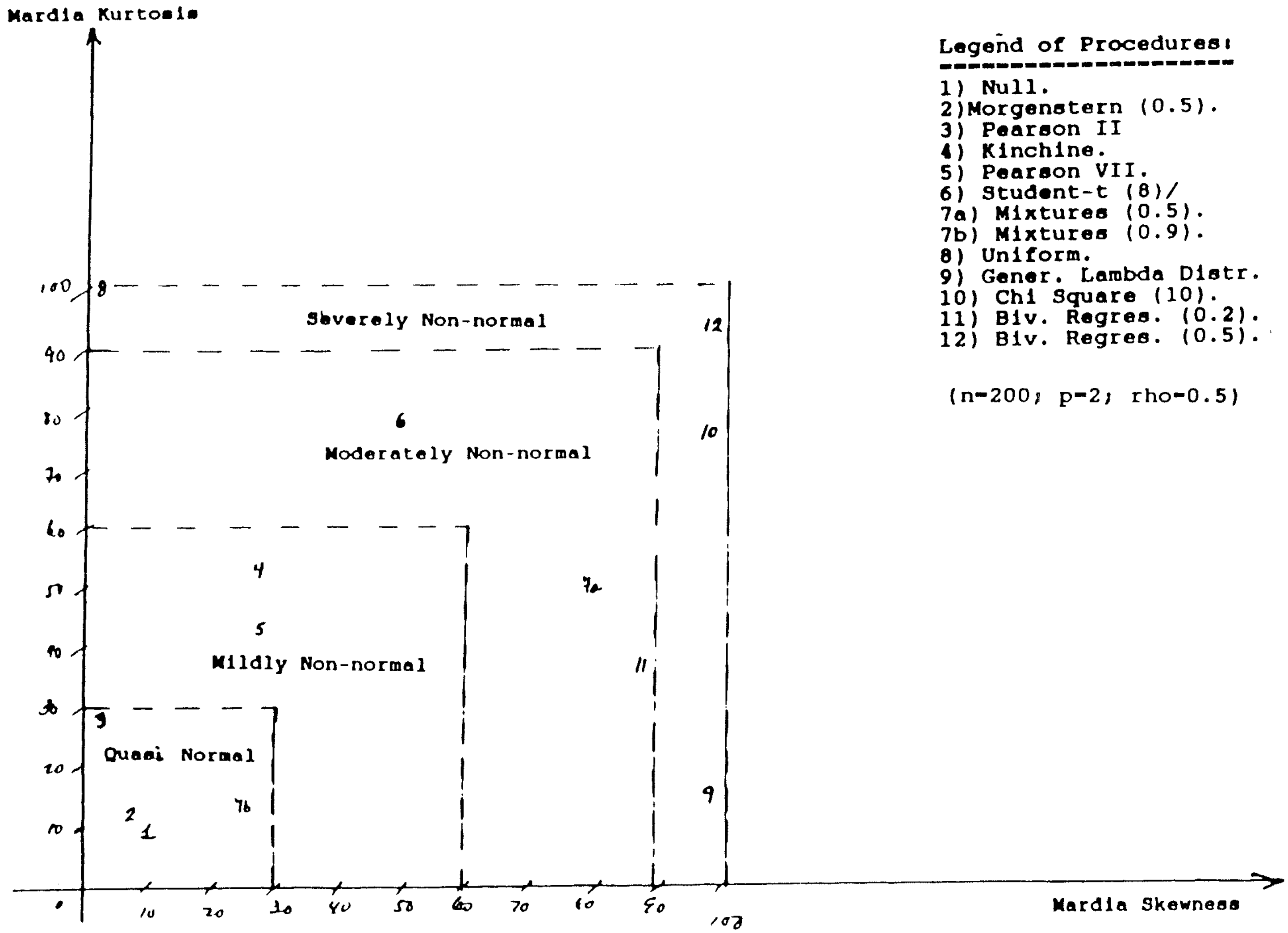


Figure 3. Achieved Power of Mardia's Skewness/Kurtosis tests, by Procedures.

tributions, given a covariance structure. But they do not place other constrains on the theoretical properties of the resulting (unknown) multivariate distribution. Such methods are easy to implement and have been widely used.

For example, Gnanadesikan (1977), obtained bivariate correlated distributions by first generating two independent random variates $Z_1, Z_2 \sim \mathcal{F}$. From them, two other variables Y_1, Y_2 are obtained by letting $Y_1 = Z_1 \quad and \quad Y_2 = \rho Z_1 + \sqrt{1-\rho^2} Z_2$ with $Corr\{Y_1, Y_2\} = \rho$. Or, as performed by Loh (1986), following Andrews et al. (1973), by applying a transformation g to each coordinate of a bivariate normal.

Some advantages of these combinations of distributions are their simplicity and realistic flavor. Their major disadvantage consists in poor control of some parameters: skewness, kurtosis and marginal variances. And also, that the resulting multivariate distribution are unknown, except in the case that the original $Z_i \sim N(\mu, \sigma)$ and $g = I$.

An alternative is to generate the random variates from an empirical family of distributions. Shapiro and Gross (1981) list criteria that empirical families should meet: (i) easy to select and (ii) to generate, and to (ii) include as wide a variety of shapes as possible. Shapiro and Gross also classify the distributions exclusively based on their third and fourth moments, $\sqrt{\beta_1}$ and β_2. Empirical families allow us to control these moments with ease.

Three widely used univariate empirical families are (a) the Generalized Lambda Distribution (GLD), (b) Johnson's Family and (c) Pearson's Family. The GLD family was originally developed for Monte Carlo studies. It is based on p, a percentile of the distribution $\mathcal{F}$:

$$x_p = \mathcal{F}^{-1}(p), \quad 0 \leq p \leq 1$$

The GLD family is defined in terms of these percentile functions by:

$$x_p = R(p) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2}, \quad 0 \leq p \leq 1$$

$$f(x) = \frac{1}{R'(p)} = \frac{\lambda_2}{\lambda_3 p^{\lambda_3 - 1} + \lambda_4 p^{\lambda_4 - 1}}, \quad 0 \leq p \leq 1$$

Tables for the four lambda parameters of the GLD, for given values of $\sqrt{\beta_1}, \beta_2$, are available (Ramberg, Dudewicz, Tadikmalla and Mykytka (1979)). The GLD allows the exploration of the effect of a change in skewness, given a fixed kurtosis or viceversa, with relative ease.

Johnson's system is based on the following transformation: $z = \gamma + \eta\kappa_j(x; \xi, \lambda)$, $j = 1, 2, 3$ (Shapiro and Gross (1981)), where $z \sim N(0, 1)$, where $\gamma, \eta, \xi, \lambda$ are parameters and where $\kappa_j(x; \xi, \lambda)$, $j = 1, 2, 3$ are three functional forms, each defining one of the three subfamilies in the system. Johnson's system, partitions the β_1 vs β_2 plane into two non overlapping regions: S_U, S_B, separated by S_L, the family of the Lognormal distributions.

Pearson's families of distribution (Kendall and Stuart (1966)) are defined by the equation:

$$\frac{df}{dx} = \frac{(x-a)f}{b_0 + b_1 x + b_2 x^2}, \quad b_i \in \mathcal{R},\ i = 0, 1, 2$$

where f is the density of the random variable X. Pearson defines seven family types. For example, his Type II is the Beta and Type III, the Gamma distribution.

The main advantage in using empirical families of distributions consists in the larger control we have on the distribution's parameters. One serious disadvantage is their restricted domain, resulting in somewhat artificial distributions.

Since the resulting multivariate distributions obtained from such combinations of univariates are not known, we called this approach the *indirect* approach. However, we can check for the *known* covariance structure and skewness/kurtosis. In Romeu (1990) we generated combinations of **GLD** to obtain experimentally required skewness. We used these prespecified values as validation parameters with which to check our results.

We can, similarly, achieve a prespecified covariance structure with *mixtures* of MVN distributions. The resulting unknown multivariate distributions help assess the effect of data contamination on power. Let $\mathbf{X} \sim \mathcal{F}$, where:

$$\mathcal{F} = p_0 MVN_p(\mu_1, \Sigma_1) + (1 - p_0) MVN_p(\mu_2, \Sigma_2)$$

$$\mathcal{C}ov(\mathbf{X}) = p_0\Sigma_1 + (1 - p_0)\Sigma_2 + p_0(1 - p_0)(\mu_1 - \mu_2)(\mu_1 - \mu_2)'$$

There are many possible combinations formed by varying the parameters given by vector μ_i, covariance matrix Σ_i, for $i = 1, 2$, and the mixing parameter p_0. Based on the graphical study by Johnson (1987), based on bivariate mixtures, and seeking a mildly versus a severely contaminated alternative, Romeu (1990) selected $\mu_1' = (0, \ldots, 0)$ and $\mu_2' = (1, \ldots, 1)$, $\rho_1 = 0.5$ and $\rho_2 = 0.9$ and covariance matrices as:

$$\Sigma_i = \begin{pmatrix} 1 & \rho_i & \ldots & \rho_i \\ \ldots & \ldots & \ldots & \ldots \\ \rho_i & \ldots & \rho_i & 1 \end{pmatrix} \qquad i = 1, 2$$

The second approach to generating multivariate distribution, which we call *direct*, is more efficient but complex and often mathematically involved or intractable. Among the methods included in this group are (i) conditional distribution, (ii) transformation of marginals and (iii) factorization.

The *conditional* distribution approach to generating a random vector (r.v.) $\mathbf{X}$ requires, first, the derivation of the p marginal distributions X_i. Then, of the successive conditional distributions of $X_j|X_{j-1}, \dots, X_1$, for $j = 2, \dots, p$. This is not always easy or feasible. For the *transformation* approach, a function $g(Y) = \mathbf{X}$ must be found such that $\mathcal{F}(g(Y)) = \mathcal{F}(\mathbf{X})$. Then, we proceed by generating, first, the easier multivariate Y. Then transforming it to $\mathbf{X}$ via g. The problem with this approach is that function g is not always available. For details, see Johnson, Wang and Ramberg (1984).

A frequent application of the above technique is in the generation of $MVN_p(\mu, \Sigma)$, form $MVN_p(0, I)$, via a Cholesky factorization A of the required covariance matrix $AA' = \Sigma$. Then, $g(Y) = AY + \mu$.

The multivariate Johnson (transformation) system (Johnson, 1987) offers the possibility of specifying many *controlled* multivariate distributions. But their derivation becomes mathematically involved and often intractable as p increases.

As in Johnson's univariate system, mentioned above, one of the four established transformations is performed on each of the p marginals. Then, the resulting joint multivariate distribution is obtained. Johnson has derived the densities of the transformed bivariate distributions. He has obtained relational functions between the original and resulting parameters and distributional moments, and has graphed, the bivariate distributions obtained with such transformations. They allow the study of specific types/levels of departures from the null, in a controlled environment. But for $p > 2$ the derivations become mathematically involved.

A comparison of the bivariate contours from Johnson's multivariate system with those obtained by mixtures of multivariate normals, appear on pages 64 to 82 and 56 to 51, respectively, in Johnson (1987). One notices how, with a convenient combination of the mixture parameters, similar statistical alternatives can be obtained. However, one ends up with with less information, using this simpler method. We opted for this second approach, in Romeu (1990), to generate some of our skewed distributions.

The third approach, which we have called *factorization*, obtains a multivariate r.v. by multiplication of two other ones via the Elliptically Contoured (EC) distributions (Johnson (1987)). EC are defined in terms of the subclass of spherically symmetrical distributions. A p dimensional $\mathbf{X} \sim \mathcal{F}$ is spherically symmetrical if $\mathcal{F}(X) = \mathcal{F}(PX)$, for all $p \times p$ orthogonal matrices P. Geometrically speaking, spherically symmetrical distributions are invariant under rotations and include the normal, t and the symmetrical cases of the Pearson, Johnson and GLD families.

We say (and denote) $\mathbf{X} \sim EC_p(\mu, \Sigma; g)$ if its density:

$$f(x) = \kappa_p |\Sigma|^{-1/2} g((\mathbf{X}_i - \mu)' \Sigma^{-1} (\mathbf{X}_i - \mu))$$

where κ_p is a normalizing constant and $g(.)$ a continous variable.

Therefore, $\mathbf{X}$ can be generated by multiplying R by $U^{(p)}$:

$$\mathbf{X} = RBU^{(p)} + \mu$$

where R is a positive random variable, independent of $U^{(p)}$, having the distribution of $\sqrt{(\mathbf{X}_i - \mu)' \Sigma^{-1} (\mathbf{X}_i - \mu)}$. And B is a $p \times p$ matrix such that $BB' = \Sigma$. Finally, $U^{(p)}$ is a random vector uniformly distributed on the unit hypersphere. Since $U^{(p)}$ is always the same, R^2 is the driver of the distribution of $\mathbf{X}$.

The univariate R^2 has density:

$$h(z) = \frac{\pi^{p/2}}{\Gamma(p/2)} \kappa_p \, z^{p/2-1} \, g(z)$$

$$where \quad z = (\mathbf{X} - \mu)' \Sigma^{-1} (\mathbf{X} - \mu)$$

For the multivariate normal (p), R^2 is the χ^2_p and $\kappa_p = (2\pi)^{-p/2}$ and g the identity. For the Pearson Type II, R^2 is $Beta(p/2, m+1)$. And for Pearson Type VII, R^2 is the univariate Pearson Type VI. This last type is generated via:

$$R^2 = Y/(1-Y), \qquad where \qquad Y \sim Beta(p/2, m - 1/2)$$

In Romeu (1990) we selected two elliptically contoured distributions: Pearson's type II and VII (Johnson (1987); Chmielewski (1981)), with parameters $m = 10, 6, 2$. Both these distributions are close to being multivariate and marginally normal, for large p.

The density function of the p-dimensional Pearson II distribution is:

$$f(x) = \frac{\Gamma(\frac{p}{2}+m+1)}{\Gamma(m+1)\pi^{\frac{p}{2}}}|\Sigma|^{-\frac{1}{2}}\{1-(\mathbf{X}_i-\mu)'\Sigma^{-1}(\mathbf{X}_i-\mu)\}^m$$

Its marginals are also Pearson type II distribution, with kurtosis:

$$\frac{3(m+\frac{p}{2}+1)}{m+\frac{p}{2}+2} \rightarrow 3 \quad as \quad m,p\rightarrow\infty$$

For $p=4$ and $p=8$, kurtosis is approximately 2.8, close to the Normal kurtosis of 3.0.

The density function for the p-dimensional Pearson type VII is:

$$f(x) = \frac{\Gamma(m)}{\Gamma(m-\frac{p}{2})}|\Sigma|^{-\frac{1}{2}}\{1+(\mathbf{X}_i-\mu)'\Sigma^{-1}(\mathbf{X}_i-\mu)\}^{-m}$$

Its marginal distribution is, again, Pearson VII. We used the property that marginals of EC distributions are also EC (in our case Pearson distributed) as another validation parameter in our Monte Carlo study.

A more recent approach to this problem is that of Rangaswamy, Weiner and Ozturk (1992). They decompose the multivariate $\mathbf{X} \sim \mathcal{F}$ using the factorization $\mathbf{X} = SZ$. Here, $Z \sim MVN_p(0,\Sigma)$ and S is a univariate r.v. driving the multivariate distribution of $\mathbf{X}$ such that:

$$f_{\mathbf{X}}(X) = (2\pi)^{-p/2}|M|^{-1/2}h_p(q); \qquad where \quad q=\mathbf{X}'M^{-1}\mathbf{X}$$

$$where \qquad h_p(q) = \int_0^\infty s^{-p}exp(\frac{-q}{2s^2})f_S(s)ds \quad and \quad \Sigma = M\mathcal{E}(S^2)$$

One can make $M=\Sigma$ by redefining $s' = \frac{s}{\mathcal{E}(s)}$. And one can obtain a general multivariate r.v. Y, by letting:

$$Y = A\mathbf{X}+b \quad where \quad AA'=\Sigma; \quad b \quad a \quad vector$$

For example, for the case of multivariate Laplace, S is distributed as a Rayleigh.

In Romeu (1992b) a two-phase approach was used to validate such generator of a multivariate Laplace. First, with $p=1$, let $z \sim N(0,1)$ and $w \sim exp(1)$. Making the transformation $y=\sqrt{2w}$ we obtain a Rayleigh distribution, with $\mathcal{E}(y^2)=2$. Redefining $s=\frac{y}{\sqrt{2}}=\sqrt{w}$ we obtain a more convenient r.v. with expectation $\mathcal{E}(s)=1$. This yields $M=\Sigma$ and the density of S is now:

$$f_S(s) = 2s\,exp(-s^2)$$

To obtain the distribution of the quadratic form $q = x'x = x^2$ Romeu finds, for $p = 1$ and following Rangaswami et al., that:

$$h_p(q) = \int_0^\infty s^{-p} exp(\frac{-q}{2s^2}) f_S(s) ds = \sqrt{\pi} exp(-\sqrt{2q})$$

And the distribution of the quadratic form q becomes:

$$f_Q(q) = \frac{1}{\sqrt{2q}} exp(-\sqrt{2q})$$

Then, by making the transformation $t = \sqrt{2q}$ Romeu obtains that $t \sim exp(1)$. Hence, for $x = sz$, with s, z and $h_p(.)$ as above with $p = 1$ we have:

$$f_X(x) = \sqrt{2\pi} |\Sigma|^{-1/2} h_p(q) = \frac{1}{\sqrt{2}} exp(-\sqrt{2}|x|)$$

Again making the transformation $u = \sqrt{2}|x|$ Romeu obtains a Double Exponential, which is a special case of the Laplace Distribution with parameter $(\lambda = \sqrt{2})$.

Samples of size $n = 25, 50, 100, 200$ from the univariate x factored as above $(x = sz)$ were generated. And the variables x, s, q were then tested for the goodness-of-fits for, respectively, the double exponential, Rayleigh and exponential distributions. Then, in phase II we generated bivariate Laplace and, using the marginal property of EC, tested the two components of $\mathbf{X}$ and the quadratic form q. It is worth noticing that, as we increase p the resulting density functions become mathematically involved, with embedded Bessel Functions. In such cases, obtaining integrable closed forms, for the Distribution Function (CDF), manageable in a simulation program, becomes difficult.

Finally, some remarks about random number generators (RNG). In our Monte Carlo studies, we used the (IMSL) routine DRNUN, for generating the Uniform. IMSL has six different variations of its multiplicative congruential uniform generator (with or without shuffling). Each new pseudorandom uniform variate x_i, $i = 1, \ldots, n$, is generated by:

$$x_i = c_j x_{i-1} \, mod(2^{31} - 1), \quad i = 1, \ldots, n$$

Shuffling uses a table of 128 uniform random variates from which the subsequent stream of uniform variates is randomly taken/replaced to prevent any possible autocorrelation in the pseudorandom variate stream. Empirical results by Nance and Overstreet suggest that shuffling does lead to an improvement (Bratley et al. (1983)) of pseudorandom generators. On the other hand, Dudewicz and Ralley (1981) show this is not always the case.

For a comprehensive discussion of pseudorandom uniform generators or pseudorandom generators at large, refer to Bratley, Fox and Schrage (1983). For a comparison of their statistical performances and for source codes, refer to Dudewicz and Ralley (1981).

As implemented in Romeu (1990, 1992), the generation of elliptically contoured distributions requires, first, the generation of the random vector $U^{(p)}$, uniformly distributed in the unit hypersphere (where the uniform distribution induces equally likely directions).

There are several schemes for obtaining such distribution (Johnson (1987)) when $p = 2, 3, 4$, and some of these schemes can be extended to cases of $p > 4$. However, the most general method, which we used extensively, is to generate p *iid* univariate standard normal variates and, then, to define:

$$U_j = \frac{X_j}{\sqrt{X_1^2 + \cdots + X_p^2}}, \quad j = 1, \ldots, p$$

The above discussed multivariate statistical distributions fulfill two important functions. First, they yield the main validation parameters we are seeking. Then, they cover the essential needs of generating alternatives in a multivariate Monte Carlo study, i.e.:

(1) Shape (as provided by skewness, kurtosis, contamination, etc.).
(2) Covariance structure (as provided by the respective ρ in the covariance matrix (low, medium on high).
(3) Sample size (small, medium and large).
(4) Number of p-variates (few, moderate and large).
(5) Significance level (of the test statistic nominal α) corresponding to the tail values say of ten, five and one percent points.

With these generators we can assess the effects of a distribution that is skewed, instead of symmetrical; peaked or flatter than the standard kurtosis; both, skewed and peaked or flat, or suffer from data contamination.

In the initial stage of a Monte Carlo study, the requirements for the experimental design, the alternative distributions to be used, and the particular generation methods are selected. Hence, this is the proper stage to define the validation parameters and the validation roadmap for the two subsequent stages.

3.0 Concurrent Stage

The second validation phase starts with the computer implementation of the study and is carried out, as a control tool, throughout the entire experiment. It consists in verifying, along the different generation stages, that the validation parameters identified in Phase I are met. If one validation parameter fails to be met, a halt and thorough investigation of the causes should be undertaken. After verification and correction, the Monte Carlo study can proceed.

The first activity of this second stage concerns the verification of any and all off-the-shelf software to be used. Several unpleasant surprises have encouraged us to (i) program many of our routines (Press et al. (1986)) and to (i) never trust any canned software.

For example, to check the validity and accuracy of the IMSL uniform RNG option 2 (URN04 in Dudewicz and Ralley (1981)), Romeu (1990) first selected a better though slower one, implementable in our machine: URN14. We ran two non-normal alternatives that depended heavily on the generation of $U(0,1)$ random variates. We verified that the results, obtained with IMSL and URN14 RNG's were very close (tables of differences are given in Romeu (1990)). Differences between empirical powers obtained from each generator, were overwhelmingly less than 0.01. Hypothesis tests on these differences were not statistically significant.

By following the validation roadmap drafted during Phase I, we can compare incomming results with those available from previous work in the literature. In Romeu (1990) we verified how our results agreed with those obtained by Mardia and Koziol, for special cases of the p-values.

Another validation procedure consisted in checking the results under the null. For, under this hypothesis, the empirical powers should match the nominal significance level α. In Romeu (1990), three standard deviations about the empirical powers, obtained using $\sqrt{\frac{\alpha(1-\alpha)}{n}}$, were used to test that they were distributed about the nominal α.

Next, all distribution properties identified in Phase I were graphically checked. Bivariate measures of covariance, skewness and kurtosis were especially usefull. We plotted bivariate densities and graphed bivariate contours from the data generated from our distribution simulators (Figures 4 and 5). And we compared them with those appearing in the studies by Johnson (1987) to verify we had achieved accuracy and the effect desired in them.

Bivariate Kinchine: alpha = 1.0.

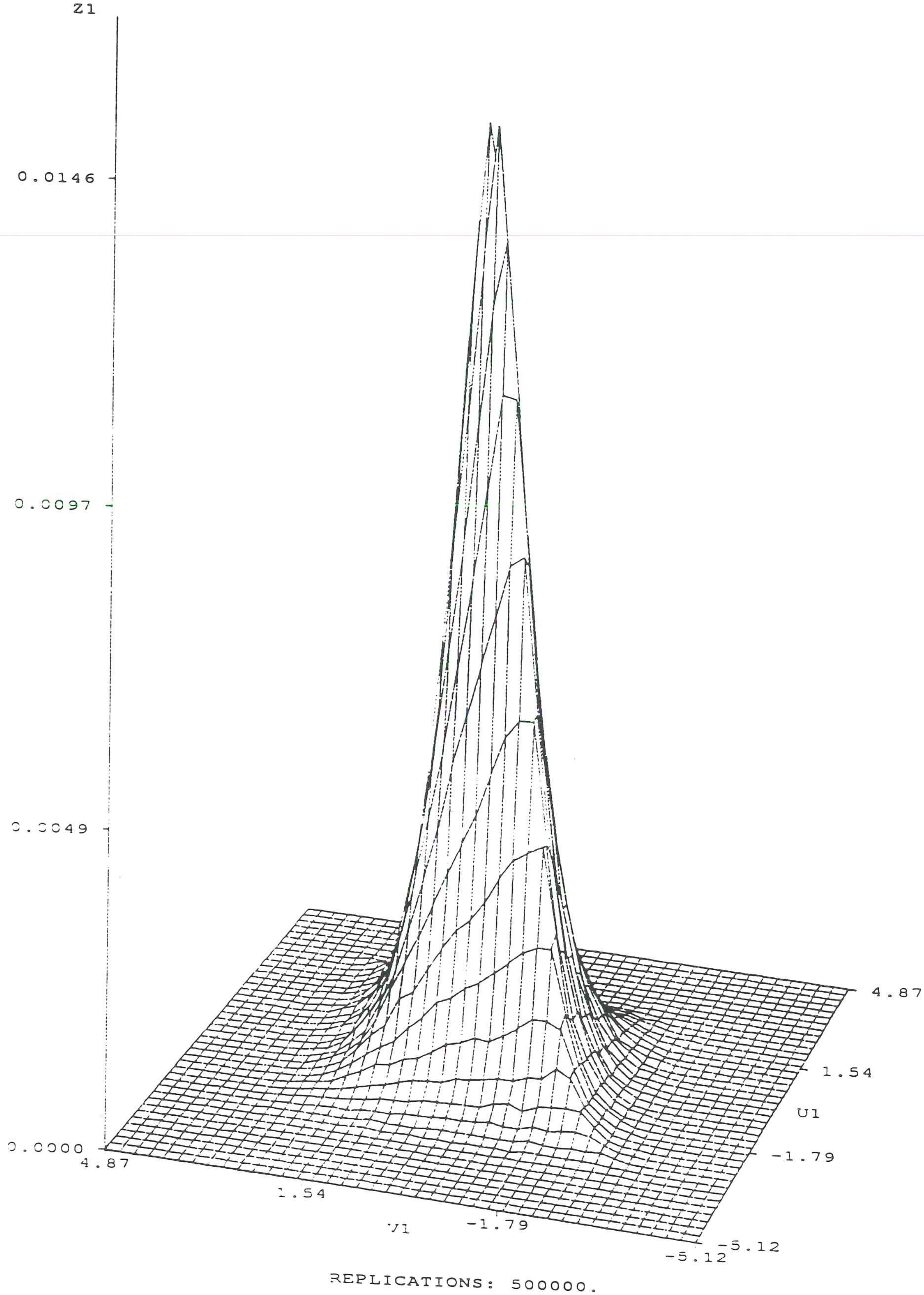


FIGURE 4.

Bivariate Kinchine: alpha = 1.0.

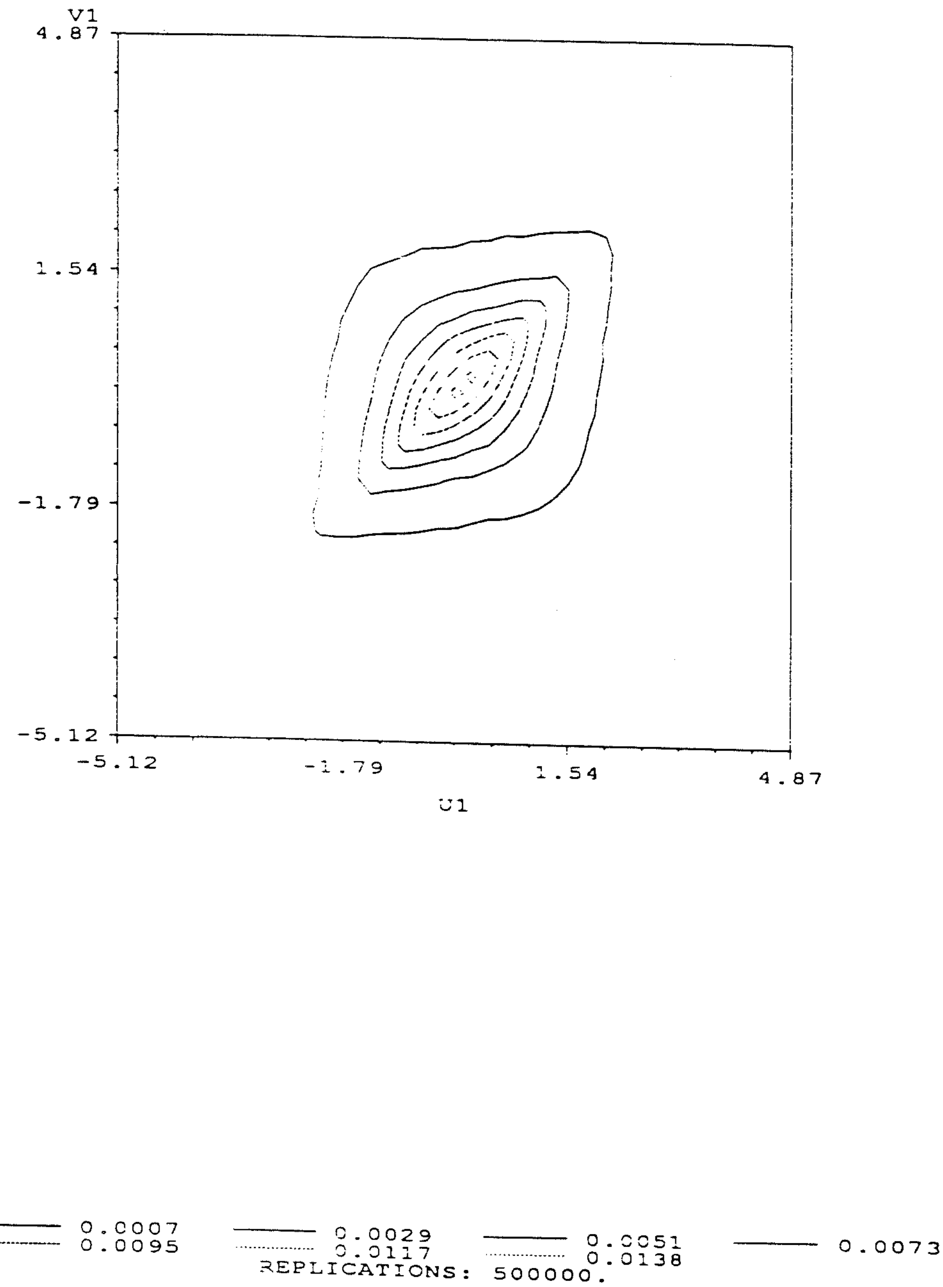


FIGURE 5.

We also used Mardia's Skewness and Kurtosis MVN GOF tests plot described in the previous section (Figure 3), to assess all generated distributions. And we verified whether or not, the positions of the different distributions were consistent with those in Figure 2. For, both graphs assessed, in different ways, the same Skewness/Kurtosis characteristics. Notice, in Figure 3, how severely skewed (GLD-1) and severely kurtic (Uniform) distributions, lay far out on the graph axes. And how severely skewed-kurtic distributions like χ^2_{10} and Bivariate Regression fall far out on the bisection of the first quadrant. And how quasi normal distributions remain about (0.1, 0.1), the nominal significance level of these tests.

Sample covariance matrices and marginal distribution parameters were also checked and compared with the theoretical ones, before launching the production runs. Sensitivity analyses were performed by letting p-variate correlation coefficient ρ vary, and comparing the results.

If there were two ways of generating the same random variable, we took advantage of this for the validation process. We implemented a small comparison to check that the results were close. For example, in Romeu (1990) we simulated the Pearson Type VII distribution of a random variable $\mathbf{X}$, via the transformation $\mathbf{X} = (\sqrt{S}/\nu)^{-1} Z + \mu$, with $Z \sim MNV_p(0, \Sigma)$ and $S \sim \chi^2_\nu$. Then we generated this distribution again, via the factorization ($\mathbf{X} = RBU^{(p)} + \mu$) approach discussed in Section 2. Results were verified and the EC method was then selected for production runs.

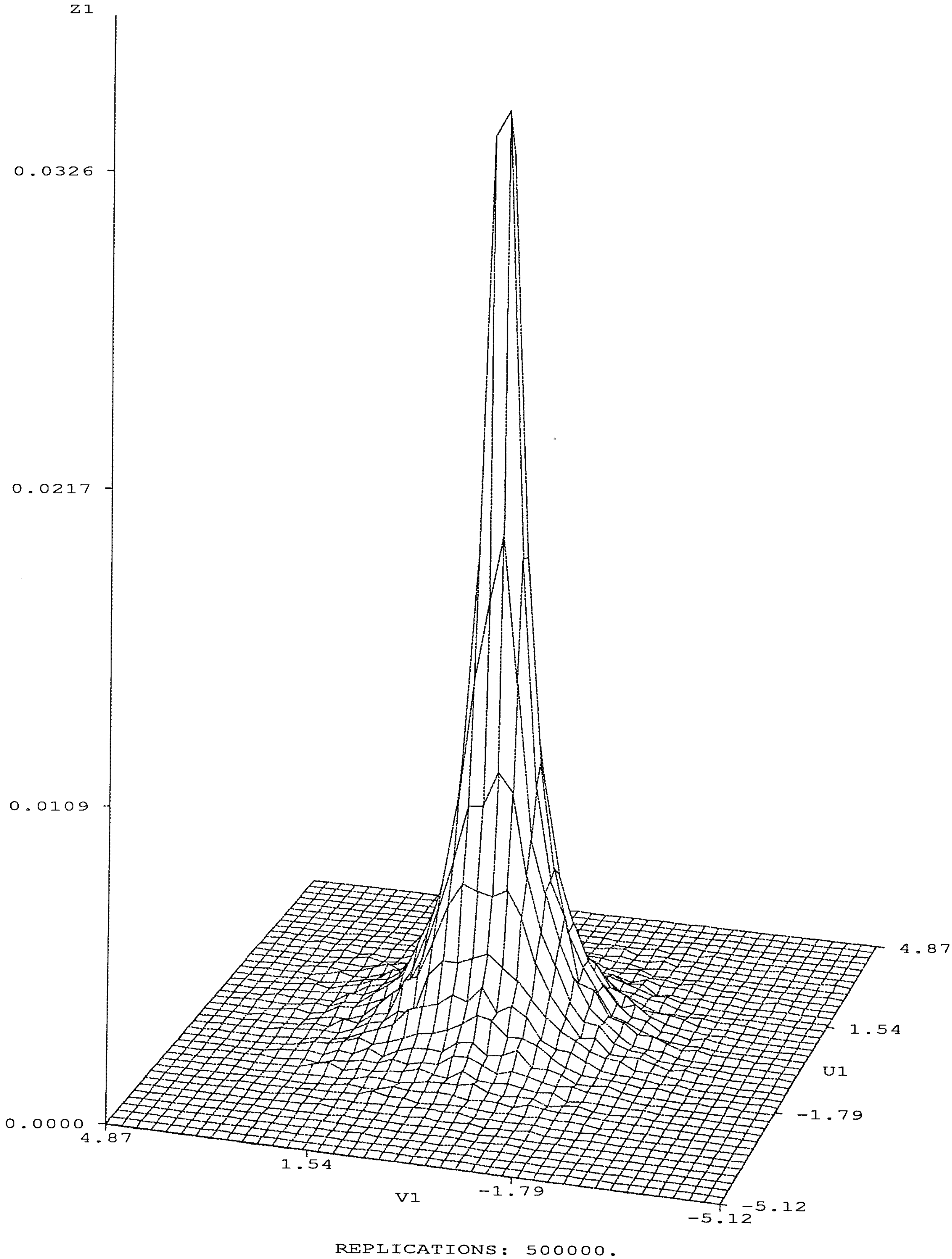
In another type of study, dealing with the validation of a theoretical generation method, Romeu (1992b) ran simulations for $n = 25, 50, 100, 200$ as explained above. We checked the fit of model variables x, s, q. We graphed the bivariate distribution, from the generated data, as a *visual validation* approach (Figure 6). We also conducted sentitivity analysis by generating closely related, but different distributions (Laplace with other values for λ). All results were positive.

Documentation of all these validation procedures and their results should be carefully kept and included in the final report with the study findings.

4.0 Final Analysis

This last validation stage is undertaken after the Monte Carlo study is completed and the final results are available. The type of validation performed in this stage is highly dependent on the objectives of the study. However, all Monte Carlo studies have in common

Figure 6. Bivariate Laplace.



that the subject problem is not tractable *in the study setting.* But that there usually exists another setting, where asymptotic results or particular cases do have a solution which is avialable for comparison.

For example, Romeu (1990) obtained small sample empirical critical values (ecv) for several MVN GOF tests, for samples $n = 25(25)200$, for $\rho = 0.5, 0.9$ and $p = 2(1)6(2)10$. Several of these tests also had asymptotic distributions. We used the asymptotic critical values as a validation parameter, by regressing the empirical ones on sample size. And then, comparing the regression independent term, β_0, with the corresponding asymptotic value.

In Table 3, we present some of these regression results. We regressed, by percentile $\eta = 0.90, 0.95, 0.99$, the ecv's on sample size n:

$$ecv_\eta = \beta_0 + \beta_1 n^{-1}$$

Results in Table 3 include, β_0, its standard deviation σ_β, the corresponding asymptotic critical value CV, and the regression index of fit R^2. As expected, as $n \to \infty$, our empirical values approached the asymptotic critical values.

Another way of validating a Monte Carlo study using the final results, in Romeu (1990), consisted of assessing the precision of the empirical critical values for $n = 200$. Following Dudewicz and van der Meulen (1984), we obtain two order statistics: $Z_{(r)}, Z_{(s)}$, such that;

$$s - 1 + 0.5 = \xi_\alpha \sqrt{Np(1-p)} + Np$$

$$r - 0.5 = -\xi_\alpha \sqrt{Np(1-p)} + Np$$

where ξ_α is the normal standard percentile evaluated at α and s, r are the corresponding positions in the ordered sample $Z_{(i)}$, $i = 1, \ldots, n$, such that:

$$\mathcal{P}\{Z_{(r)} < \xi_p < Z_{(s)}\} \geq p$$

We calculated approximate 95% conficence intervals (C.I.) for the relevant critical values. Examples of such calculations are presented in Table 4, for $p = 2, 5, 8$, $n = 200$ and $\rho = 0.5, 0.9$. There, we show, the point estimate of the 95^{th} percentile (ecv), the corresponding lower/upper bounds of its approximate 95% C.I. ($Z_{(r)}, Z_{(s)}$), and the asymptotic critical value. We can verify how these C.I. cover the asymptotic values in some MVN GOF tests

Table 3: Regressions of ecv on Sample Sizes.

ROW	p	eta	ecv	C.V.	sigma	IoF	MVN GOF Test:
1	2	0.90	7.790	7.78	0.027	0.99	Mardia Skew.
2	2	0.95	9.670	9.49	0.044	0.98	" "
3	2	0.99	13.770	13.28	0.065	0.98	" "
4	2	0.90	4.580	4.61	0.014	0.98	Cox and Small
5	2	0.95	5.990	5.99	0.032	0.94	" "
6	2	0.99	9.100	9.21	0.078	0.94	" "
7	2	0.90	0.989	*	5.500&	0.99	Malkovich-Afifi
8	2	0.95	0.989	*	5.800&	0.99	" "
9	2	0.99	0.989	*	7.500&	0.99	" "
10	2	0.90	1.630	1.65	0.012	0.99	Mardia Kurt.(LB)
11	2	0.95	2.080	1.95	0.016	0.99	" "
12	2	0.99	2.900	2.58	0.036	0.92	" "
13	2	0.90	-1.550	-1.65	0.011	0.96	" " (UB)
14	2	0.95	-1.770	-1.95	0.018	0.95	" "
15	2	0.99	-2.180	-2.58	0.032	0.93	" "
16	2	0.90	1.060	*	5.200&	0.14@	Hawkins
17	2	0.95	1.320	*	7.700&	0.04@	"
18	2	0.99	1.940	*	1.500&	0.08@	"
19	2	0.90	4.630	4.61	0.037	0.30@	Koziol Angles
20	2	0.95	5.990	5.99	0.060	0.11@	" "
21	2	0.99	9.110	9.21	0.157	0.04@	" "

* Empirical tests; no asymptotic distribution available.

@ Critical Value was independent of sample size.

& Exponential notation; four decimal places (i.e 5.5*e-4).

Table 4: 95% Nonparametric Confidence Intervals.

ROW	p	rho	eta	CV	LB	UB	MVN GOF Test:
1	2	0.5	0.95	9.49	9.17	9.660	Mardia's Skew.
2	2	0.5	0.90	1.65	1.52	1.610	" Kurt.
3	2	0.5	0.95	5.99	6.11	6.533	Royston
4	2	0.5	0.95	5.99	5.94	6.290	Cox and Small
5	2	0.5	0.95	5.99	5.85	6.110	Koziol Angles
6	2	0.9	0.95	9.49	9.11	9.590	Mardia's Skew.
7	2	0.9	0.90	1.65	1.48	1.570	" Kurt.
8	2	0.9	0.95	4.61	4.78	5.070	Royston
9	2	0.9	0.95	5.99	5.85	6.230	Cox and Small
10	2	0.9	0.95	5.99	5.86	6.220	Koziol Angles
11	5	0.5	0.95	49.80	48.32	49.710	Mardia's Skew.
12	5	0.5	0.90	1.65	1.23	1.340	" Kurt.
13	5	0.5	0.95	11.07	10.83	11.410	Royston
14	5	0.5	0.95	11.07	20.05	26.990	Koziol Angles
15	5	0.9	0.95	49.80	48.43	49.760	Mardia's Skew.
16	5	0.9	0.90	1.65	1.27	1.410	" Kurt.
17	5	0.9	0.95	5.33	8.24	8.800	Royston
18	5	0.9	0.95	11.07	18.97	22.920	Koziol Angles
19	5	0.5	0.99	57.34	56.41	58.920	Mardia's Skew.
20	5	0.5	0.98	2.05	2.02	2.300	" Kurt.
21	8	0.5	0.95	145.98	143.32	145.610	" Skew.
22	8	0.5	0.90	1.65	1.08	1.190	" Kurt.
23	8	0.5	0.95	15.51	13.53	14.480	Royston
24	8	0.5	0.95	15.51	16.25	17.090	Koziol Angles
25	8	0.9	0.95	145.98	142.14	144.750	Mardia's Skew.
26	8	0.9	0.90	1.65	1.06	1.200	" Kurt.
27	8	0.9	0.95	5.55	9.12	9.660	Royston
28	8	0.9	0.95	15.51	18.25	19.530	Koziol Angles

* Rho is the p-variate correlation coefficient.

* ecv's LB (confidence interval lower bounds) and UB (upper bounds) were empirically obtained with 10,000 replications for p=2 and with 5,000 replications for p>2.

* CV is the asymptotic critical value, for the corresponding percentile, (eta), of 90, 95, 98 or 99 percent for the test in question.

that converge faster (e.g. Mardia's Skewness test) while they do not cover (for $n = 200$) the asymptotic value in those tests that converge slower (e.g. Mardia's Kurtosis, converges for $n > 200$).

Any other similar asymptotic result or special case can be exploited for performing validation at this final stage. In Romeu (1992b), the distribution of interest was the K-Distribution, which requires the manipulation of Bessel functions. Laplace, a special case of the K, with parameter $\alpha = 1$, was used. Such special cases will surface during the literature search perform in Phase I.

5.0 Conclusions

Validation of *multivariate Monte Carlo studies* constitutes an involved, time consuming but necessary activity. It really consists of a three-phase, continous process that starts with the initial planning of the study and concludes after the study final results are obtained.

As opposed to system simulation, Monte Carlo studies do not have real data to compare with, from an operating system. As opposed to univariate statistics, multivariate introduces yet another level of complexity to the problem. For now, the distributions obtained by transformations are much more complex. They have too many parameters and factors to control.

However, some of these same characteristics can be used as a validating tool. If we have correctly performed transformations and generation of variates, the output must exhibit certain prespecified caracteristics that we want to measure. Another problem arises because, sometimes, these characteristics are difficult to measure.

In any case, the confidence in the results of any Monte Carlo study lies in the faith we can place in it. Hence, all efforts allocated to the validation process are well justified.

There is a final conclusion, that surfaces inevitably in the mind of the educator, after implementing an experience such as this. We have seen how the validation process requires extensive study and implementation of distributions, its moments, estimators, transformations, generation, hypothesis testing and experimental design, among other statistical topics.

Therefore, a *project* such as the *design, implementation and validation of* a small *multivariate Monte Carlo power study* lends itself, beautifully, for an *applied graduate statistics course*.

References

Anderson, T. W., *An Introduction to Multivariate Analysis*, Wiley, 1984.

Andrews, D. F., Gnanadesikan R. and J. L. Warner, *Methods for Assessing Multivariate Normality*, Multivariate Analysis, Academic Press, 1973.

Bratley, P; Fox, B. and L. Schrage, *A Guide To Simulation*, Springer-Verlag, 1983.

Cambanis, S.; Huang, S. and G. Simons, *On the Theory of Elliptically Contoured Distributions*, J. Multivar. Annal. **11** (1981), 368–385.

Chmielewski, M. A., *A Re-Appraisal of Tests for Normality*, Comm. Stat. - Theor. Meth. **a10(20)** (1981), 2005–2014.

Dudewicz, E. J. and T. G. Ralley, *The Handbook of Random Number Generation and Testing With TESTRAND Computer Code*, American Sciences Press Inc., 1981.

Dudewicz, E. J. and E. C. van der Meulen, *On Assessing the Precision of Simulation Estimates of Percentile Points*, Amer. Jour. Math. Manag. Sci. **4 (3-4)** (1984), 335–343.

Gnanadesikan, R., *Methods of Statistical Data Analysis of Multivariate Observations*, Wiley, 1977.

Johnson, M. E., *Multivariate Statistical Simulation*, Wiley, 1987.

Johnson, N. L. and S. Kotz, *Distributions in Statistics: Continous, Univariate Distributions*, Wiley, 1970.

Johnson, M. E., Chiang, W. and J. S. Ramberg, *Generation of Continuous Multivariate Distributions for Statistical Applications*, Amer. Jour. Math. Manag. Sci. **4** (1984), 225–248.

Johnson, R. A. and D. E. Wichern, *Applied Multivariate Analysis*, Prentice Hall, 1982.

Johnson M. and J. Ramberg, *The Johnson Translation System in Monte Carlo Studies*, Comm. Stat. - Simula. Comput. **11(5)**, 521–525.

Kendall, M. G. and A. Stuart, *The Advanced Theory of Statistics*, (Vols. I, II and III), Charles Griffin and Co., London, 1966.

Koziol, J. A., *A Class of Invariant Procedures for Assessing Multivariate Normality*, Biometrika **69** (1982), 423–427.

Koziol, J. A., *On Assessing Multivariate Normality*, JRRS-B **45** (1983), 358–361.

Koziol, J. A., *Assessing Multivariate Normality: A Compendium*, Comm. Stat. **15** (1986), 2763–2783.

Loh W., *Testing Multivariate Normality by Simulation*, Jour. Statist. Comput. Simul. **26** (1986), 243–252.

Mardia, K. V., Kent, J. T. and J. M. Bibby, *Multivariate Analysis*, Academic Press, 1979.

Mardia K. V., *Measures of Multivariate Skewness and Kurtosis With Applications*, Biometrika **57** (1970), 519–530.

Mardia, K. V., *Assessment of Multinormality and the Robustness of Hotelling T Test*, Appl. Statist. **24** (1975), 163–171.

Ozturk, A. and J. L. Romeu, *A New Graphical Test for Multivariate Normality*, Comm. in Statist. (Simula.) **21(1)** (1992).

Press, W. H.; Flannery, B. P.; Teukolsky, S. A. and W. T. Vetterling, *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, 1986.

Ramberg, J. S.; Dudewicz, E. J.; Tadikamalla, P. R. and E. F. Mykytka, *Probability Distributions and its Uses in Fitting Data*, Technometrics **21, no. 2** (1979), 201–214.

Rangaswamy, M.; Weiner, D. and A. Ozturk, *Computer Generation of Correlated Non Gaussian Clutter for Radar Signal Detection*, IEEE Trans. Aerosp. Electr. Sys. (1992 (to appear)).

Romeu, J. L., *A Simulation Approach for the Analysis and Forecast of Software Productivity*, Journal of Computers and Industrial Engineering **9(2)** (1985).

Romeu, J. L., *Teaching Engineering Statistics With Simulation: A Classroom Experience*, Journal of the Institute of Statisticians **35(4)** (1986).

Romeu, J. L., *Another Look at the Comparison of the Non Overlapping Batch Means and Area STS Simulation Output Analyses Procedures*, Actas del ISORBAC-2, San Sebastian, 1988.

Romeu, J. L., *A Small Sample Monte Carlo Study of Four System Reliability Bounds*, Journal of Computers and Industrial Engineering **16(1)** (1989).

Romeu, J. L., *Development and Evaluation of a General Procedure for Assessing Multivariate Normality*, CASE Center Technical Report 9022. Syracuse University, NY. 13244, 1990.

Romeu, J. L., *A New Multivariate Normality Goodness of Fit Test With Graphical Applications*, Proceedings of the Computers and Industrial Engineering Conference (1991).

Romeu, J. L., *Small Sample Empirical Critical Values as a Tool for the Comparison of Multivariate Normality Goodness of Fit Tests*, Proceedings of the Conference on the Interface Between Statistics and Computer Science (1992a).

Romeu, J. L., *Monte Carlo Validation of a Theoretical Model for Generating Non Gaussian Radar Clutter*, Final Report. U.S. Air Force Faculty Summer Research Program. (To Appear), 1992b.

Shapiro, S. and A. Gross, *Statistical Modeling Techniques*, Marcel Dekker, 1981.

Tong, Y. L., *The Multivariate Normal Distribution*, Springer-Verlag, 1990.