Fall 1971

A New Test for Normality

Richard LeRoy Roller
Central Washington University

A NEW TEST FOR NORMALITY

A Thesis

Presented to

the Graduate Faculty

Central Washington State College

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Richard LeRoy Roller

December, 1971

APPROVED FOR THE GRADUATE FACULTY

____

William B. Owen, COMMITTEE CHAIRMAN

____

David R. Anderson

____

Frederick M. Lister

TABLE OF CONTENTS

A NEW TEST FOR NORMALITY

by

Richard L. Roller

December, 1971

This paper presents a new test for normality which is based on a complete characterization of the normal distribution. Motivation for the test is given in terms of a proof of this characterization. The test is derived and evaluated by computer-simulated sampling from alternative distributions. The empirical powers of the test generated from such samplings are tabled and compared to nine commonly used tests. Evaluation of the proposed test is discussed and further avenues of investigation are suggested.

CHAPTER I

INTRODUCTION

Any student who has had an introductory course in statistics can cite many of the properties of the normal distribution, such as the fact that the mean, median, and mode are coincidental and the distribution is symmetric about its mean, and the fact that a graph of the distribution resembles a bell. In such a course, much time is spent on properties of the normal and applications to random variables in the physical world which closely approximate the normal. Oftentimes, students, as well as experimenters, are interested in testing hypotheses or setting confidence intervals under the assumption that the population being sampled is normal. Thus, it is quite reasonable to investigate that assumption, and to try to find some criteria for determining whether or not a given population is normally distributed.

Fortunately there are many random variables in the physical world which closely approximate the normal distribution. Heights and weights of individuals, gas mileages of automobiles, and intelligence-quotients are a few diverse examples of the many real-world variables which can be approximated by the normal distribution.

Another reason for the importance of the normal distribution is that it is the building block from which many useful sampling distributions are formed. Examples of these distributions are the Chi-square with n degrees of freedom [$\chi^2(n)$], "Student's" t with n degrees of freedom [$t(n)$], Snedecor's F with n, m degrees of freedom [$F(n,m)$], the gamma distributions [$\Gamma(n,m)$], and the family of beta distributions [$\beta(n,m)$].

The preceding discussion suggests that the normal distribution is quite important in sampling theory and that we would benefit greatly if we had some way to test a distribution for normality. In Chapter II, various tests are examined that have been proposed for normality. In Chapter III, the normal density is characterized as the only density for which the sample mean and variance are independent under random sampling. This characterization is used to derive a new test for normality in Chapter IV. The power of the new test is derived in Chapter V, and in Chapter VI, the proposed test is compared with nine standard tests. The remaining chapters are devoted to analyzing weaknesses of the test and proposing avenues of improving the test.

CHAPTER II

PRESENT TESTS FOR NORMALITY

The need for determining normality has given rise to many tests. A comprehensive comparison and evaluation of the most widely used of these tests was given by Shapiro and Wilk. [2] This study provides a basis for comparison of the normality test proposed in this paper with existing tests. Following is the list of the tests compared, their code names, and a description of each test.

Shapiro and Wilk (W).

$$W = \left[\sum_{i=1}^{[n/2]} a_{n-i+1} [Y_{n-i+1} - Y_i] \right]^2 \Big/ \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

where [n/2] = greatest integer in n/2 and

a_{n-i+1} = coefficients tabulated by Shapiro and Wilk.

Standard Third Moment ($\sqrt{b_1}$).

$$\sqrt{b_1} = \sqrt{n} \sum_{i=1}^{n} (Y_i-\overline{Y})^3 \Bigg/ \left[\sum_{i=1}^{n} (Y_i-\overline{Y})^2 \right]^{3/2}$$

Standard Fourth Moment (b_2).

$$b_2 = n \sum_{i=1}^{n} (Y_i-\overline{Y})^4 \Bigg/ \left[\sum_{i=1}^{n} (Y_i-\overline{Y})^2 \right]^{2}$$

Kolmogorov-Smirnov (KS).

$$KS = \max \left| i/n - F(Y_i) \right|, \; i = 1,2,\ldots,n$$

where F is the hypothesized normal cumulative distribution function.

Cramer-Von Mises (CM).

$$CM = n \int_0^1 [F_n(Y) - F(Y)]^2 \, df(Y)$$

where F_n is the empirical distribution function.

Weighted Cramer-Von Mises (WCM).

$$WCM = n \int_0^1 [F_n(Y) - F(Y)]^2 \frac{dF(Y)}{F(Y)[1-F(Y)]}$$

Durbin's Modified KS (D).

$$D = \max_i \left[\frac{i}{n} - \sum_{j=1}^{i} g_j \right], \quad i=1,2,\ldots,n$$

where $g_j = (n+2-j)(c_j^* - c_{j-1}^*), \quad j=1,2,\ldots,n,$

$0 \leq c_0^* \leq c_1^* \ldots \leq c_n^*$ obtained by ordering

$c_1 = u_1, c_2 = u_2 - u_1, \ldots, c_{n+1} = 1 - u_n$ and

$u_i = F(Y_i)$, $i=1,2,\ldots n$.

Chi-squared (Equiprobable Cells) (CS).

$$CS = \frac{k}{n} \sum_{i=1}^{k} c_i^2 - n$$

where k=number of cells and c_i= number of observations per cell.

David et al (U).

$$U = \sqrt{(n-1)}\,(Y_n - Y_1) \Big/ \left[\sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right]^{1/2}$$

where $Y_1 \leq Y_2 \leq \ldots \leq Y_n$.

The results of the Shapiro-Wilk-Chen article showed that the W statistic displayed consistently more sensitivity than did any other test. The distance tests (KS, CM, WCM, and D) proved to be typically very insensitive, while the U statistic proved quite powerful against short-tailed symmetric distributions but had no sensitivity to asymmetry. A combination of $\sqrt{b_1}$ and b_2 was found to have good power but even this combination was dominated by W.

The results were simulated on a high speed computer, and twelve families of alternative distributions were sampled, with sample sizes ranging from ten to fifty. Some of the results will be presented later for comparison.

CHAPTER III

CHARACTERIZING THE NORMAL DENSITY

It has long been known that the normal distribution is characterized by the fact that the sample mean and sample variance are statistically independent. Hence it would seem reasonable to base a test for normality on this fact. Furthermore, a test based on this complete characterization would intuitively seem to be quite powerful. None of the nine tests considered in the article, or for that matter any test known to this writer, has used this characterization. Following is a proof of this characterization. [1:362-364]

Theorem. $\overline{X}$ and S^2 are independent if and only if the X_i are normally and identically distributed random variables with mean μ and variance σ^2, where $\overline{X} = \Sigma X_i/n$ and $S^2 = 1/n\Sigma X_i{}^2 - \overline{X}^2$.

Proof. Suppose that we have a population with finite mean μ and variance σ^2 and characteristic function $\phi(t)$. The joint characteristic function for the sample mean and variance taken from a random sample of size n is given by

$$\phi_{12}(t_1,t_2) = \int_{\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(it_1\bar{X} + it_2S^2)\, dF_n.$$

A necessary and sufficient condition that $\bar{X}$ and S^2 be independent is that the joint characteristic function factors into the product of the marginal characteristic functions or

$$\phi_{12}(t_1,t_2) = \phi_1(t_1)\ \phi_2(t_2).$$

Taking partials of the above equation with respect to t_2 we get

$$\left[\frac{\partial\phi_{12}}{\partial\ t_2}\right]_{t_2=0} = \phi_1(t_1)\left[\frac{\partial\phi_2}{\partial t_2}\right]_{t_2=0} \qquad (3.1)$$

Note that

$$\phi_1(t_1) = \int\cdots\int \exp(it_1\bar{X})\,dF_n = \int\cdots\int \exp\left[(it_1/n)\Sigma X_i\right]dF_n$$

$$= \int\cdots\int \exp\left[(it_1/n)X_1\right]\cdots \exp\left[(it_1/n)X_n\right]\, dF_n$$

$$= \int \exp\left(\frac{it_1}{n} X_1\right) dX_1 \int \exp\left(\frac{it_1}{n} X_2\right) dX_2 \cdots \int \exp\left(\frac{it_1}{n} X_n\right) dX_n$$

$$= \phi(t_1/n) \ \ldots \ \phi(t_1/n).$$

Therefore, $\phi_1(t_1) = [\phi(t_1/n)]^n$.

Therefore, we may write (3.1) as

$$\left[\frac{\partial \phi_{12}}{\partial t_2}\right]_{t_2=0} = [\phi(t_1/n)]^n \left[\frac{\partial \phi_2}{\partial t_2}\right]_{t_2=0}.$$

Now consider

$$\phi_2(t_2) = \int \cdots \int \exp(it_2 S^2) \ dF_n$$

$$\frac{\partial \phi_2}{\partial t_2} = i \int \cdots \int S^2 \exp(it_2 S^2) \ dF_n$$

$$\left[\frac{\partial \phi_2}{\partial t_2}\right]_{t_2=0} = i \int \cdots \int S^2 \exp(0) \ dF_n$$

$$= i \int \cdots \int S^2 \ dF_n$$

$$= i \ E(S^2)$$

$$= i \ E\{(1/n)\Sigma X_i^2 - \overline{X}^2\}$$

$$= \frac{i(n-1)\sigma^2}{n}.$$

Substituting this into (3.1) gives

$$\left[\frac{\partial\phi_{12}}{\partial t_2}\right]_{t_2=0} = \left[\phi\left(\frac{t_1}{n}\right)\right]^n \frac{i(n-1)\sigma^2}{n} .$$

Taking the partial of the joint characteristic function with respect to t_2 yields

$$\frac{\partial\phi_{12}}{\partial t_2} = i\int\cdots\int S^2\exp(it_1\overline{X} + it_2S^2)\, dF_n,$$

and since $S^2 = 1/n\Sigma X_i^{\ 2} - \overline{X}^2 = \frac{1}{n^2}\left((n-1)\Sigma X_i^{\ 2} - \sum_{i\neq j}\sum X_iX_j\right)$,

the above equation becomes

$$\frac{\partial\phi_{12}}{\partial t_2} = \frac{i}{n^2}\int\cdots\int\left((n-1)\Sigma X_i^{\ 2} - \sum_{i\neq j}\sum X_iX_j\right)\exp(it_1\frac{\Sigma X_i}{n} + it_2S^2)\, dF_n.$$

At t_2=0 this reduces to

$$\left[\frac{\partial\phi_{12}}{\partial t_2}\right]_{t_2=0} = \frac{i}{n^2}\int\cdots\int\left((n-1)\Sigma X_i^{\ 2} - \sum_{i\neq j}\sum X_iX_j\right)\exp\left(it_1\frac{\Sigma X_i}{n}\right) dF_n.$$

Now, since the X_i are mutually independent,

$$\left[\frac{\partial\phi_{12}}{\partial t_2}\right]_{t_2=0} = \frac{i}{n^2}\left\{n(n-1)\int\cdots\int X^2\exp\left(i\,\frac{t_1}{n}\,X\right)dF_n\right.$$

$$\left.- n(n-1)\left[\int\cdots\int X\exp\left(i\,\frac{t_1}{n}\,X\right)dF_n\right]^2\right\}$$

$$= \frac{i(n-1)}{n}\left\{\int X^2\exp\left(i\,\frac{t_1}{n}\,X\right)dF_1\left[\phi\left(\frac{t_1}{n}\right)\right]^{n-1}\right.$$

$$\left.-\left[\int X\exp\left(i\,\frac{t_1}{n}\,X\right)dF_1\right]^2\left[\phi\left(\frac{t_1}{n}\right)\right]^{n-2}\right\}$$

$$= \frac{i(n-1)}{n}\left[\phi\left(\frac{t_1}{n}\right)\right]^{n-2}\left\{-\phi\left(\frac{t_1}{n}\right)\frac{\partial^2\phi\left(\frac{t_1}{n_1}\right)}{\partial\left(\frac{t_1}{n}\right)^2} + \left[\frac{\partial\phi\left(\frac{t_1}{n}\right)}{\partial\left(\frac{t_1}{n}\right)}\right]^2\right\}.$$

Hence, (3.1) becomes

$$\frac{i(n-1)}{n}\left[\phi\left(\frac{t_1}{n}\right)\right]^{n-2}\left\{-\phi\left(\frac{t_1}{n}\right)\frac{\partial^2\phi\left(\frac{t_1}{n}\right)}{\partial\left(\frac{t_1}{n}\right)^2} + \left[\frac{\partial\phi\left(\frac{t_1}{n}\right)}{\partial\left(\frac{t_1}{n}\right)}\right]^2\right\}$$

$$= \left[\phi\left(\frac{t_1}{n}\right)\right]^n \frac{i(n-1)\sigma^2}{n}.$$

Letting $t = \frac{t_1}{n}$, and rearranging, we have

$$-\phi(t)\ \phi''(t) + [\phi'(t)]^2 = [\phi(t)]^2\sigma^2 .$$

But this may be written as

$$\frac{d}{dt}\left\{\frac{d}{dt}\ \log\phi(t)\right\} = -\sigma^2.$$

Integrating with respect to t yields

$$\frac{d}{dt}\ \log\phi(t) = -\sigma^2 t + C_1.$$

But $\phi(0) = \int \exp(i\cdot 0\cdot x)\ dF_1 = 1$ and

$$\phi'(0) = i\mu\int \exp(i\cdot 0\cdot x)\ dF_1 = i\mu,$$

which gives $C_1 = i\mu$.

Hence,

$$\frac{d}{dt}\ \log\phi(t) = -\sigma^2 t + i\mu,$$

and integrating again with respect to t yields

$$\log\phi(t) = \frac{-\sigma^2 t^2}{2} + i\mu t + C_2.$$

But $\log\phi(0) = \log(1) = 0$, which says $C_2 = 0$, and we finally have

$$\phi(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2} \right),$$

which is the unique characteristic function of the univariate normal among all distributions with finite variances.

CHAPTER IV

DERIVING A NEW TEST FOR NORMALITY

Because of the characterization of the normal just proved, the problem of testing for normality is reduced to testing for independence of $\overline{X}$ and S^2.

Consider the following scheme: take a random sample of size 4n from some distribution which we wish to test for normality. Pair the observations two at a time, forming 2n pairs. Find the sample mean, $\overline{X}_i$, and the sample variance S_i^2, for each of the pairs, and then compute the median (M_1) of $\overline{X}_1, \ldots, \overline{X}_n$ and the median (M_2) of $S_1^2, \ldots, S_n^2$. The test statistic T is computed by counting the number of pairs whose mean lies above M_1 and whose variance is below M_2. (See Figure 4.1)

$\overline{X}$ \ S^2	Above M_2	Below M_2
Above M_1	N_{11}	$N_{12} = T$
Below M_1	N_{21}	N_{22}

FIGURE 4.1

Assuming independence of the two categorized variables, the expected values of the cell entries are given by

$$e_{ij} = \frac{N_{i.}\,N_{.j}}{N_{..}} \quad i,j = 1,2.$$

Clearly e_{ij} depends only on the marginal frequencies and the sample size. These marginal totals are all fixed and equal to n since with 2n values of $\overline{X}$ and S^2, there will always be n values of each above and below their respective medians. Figure 4.2 shows the completed 2x2 table of expected values assuming independence of $\overline{X}$ and S^2, and hence normality of the underlying distribution.

		S^2 Above M_2	S^2 Below M_2	
$\overline{X}$	Above M_1	n/2	n/2	n
$\overline{X}$	Below M_1	n/2	n/2	n
		n	n	2n

FIGURE 4.2

Since the marginals are fixed, only one of the cell entries in the observed data table may vary independently. Let that random variable be T, the observed number of $(\overline{X}, S^2)$ pairs that fall in the "above M_1, below M_2" cell. Then under fixed marginals the distribution of T is the complete distribution of the table. The probability distribution of T under normality can be derived in the following manner. A random sample is taken without replacement from 2n pairs of $(\overline{X}, S^2)$'s, n of which have $\overline{X}$'s above M_1 and n of which have $\overline{X}$'s below M_1. Let A(t) be the event that exactly t of the $(\overline{X}, S^2)$ pairs with $\overline{X}$ above M_1 have corresponding S^2 below M_2. Then the probability that T=t is given by the ratio of the number of possible outcomes in A(t) to the total number of possible outcomes. The number of possible outcomes in A(t) is given by

$$\binom{n}{t}\binom{n}{n-t} .$$

The total number of possible outcomes is given by $\binom{2n}{n}$, hence,

$$P[T=t] = \frac{\binom{n}{t}\binom{n}{n-t}}{\binom{2n}{n}}$$

$$= \frac{[n!]^4}{[t!]^2[(n-t)!]^2(2n)!} \quad .$$

It is clearly seen that the probability distribution of T follows the hypergeometric distribution, and hence appropriate critical regions can be determined from the hypergeometric distribution. For example, with 4n=8 observations, there are only 4 $\overline{X}$'s and 4 S^2's, and T may take only the values 0, 1, and 2. The probability distribution of T is

t	0	1	2
P[T=t]	.167	.666	.167

Note that P[T=t] is symmetric about the value 1, which is the expected value of T under independence. Thus, if we reject when T = 0 or 2, we have a test of size .334. This is much too large for practical use, and this necessitates the use of a <u>randomized</u> test. To achieve a two-tailed test of size .05 we would have to adopt the following rule: If T = 0 or 2 reject the hypothesis of independence (and thus normality) with probability .15. Thus if T = 0 or 2, a random integer between one and 100 could be selected, and if it fell between one and fifteen, the hypothesis would be rejected; otherwise, it would be accepted.

Following are the distributions of the test statistic $T = N_{12}$ for varying sample sizes; also included are appropriate randomized critical regions for the test statistic. The tables are interpreted as follows. Suppose a random sample of size 8 has been taken and a 5% test is desired. Then a value of T = 0 or 2 would give rise to rejecting the null hypothesis with probability .15. If T happened to equal one then no rejection could be made.

8 observations:

t	0	1	2
P[T=t]	.167	.666	.167
α =.001	.3%		.3%
α = .01	3%		3%
α = .02	6%		6%
α = .05	15%		15%
α = .10	30%		30%

12 observations:

t	0	1	2	3
P[T=t]	.050	.450	.450	.050
α =.001	1%			1%
α = .01	10%			10%
α = .02	20%			20%
α = .05	50%			50%
α = .10	all			all

16 observations:

t	0	1	2	3	4
P[T=t]	.014	.229	.514	.229	.014
α =.001	3.6%				3.6%
α = .01	36%				36%
α = .02	72%				72%
α = .05	all	4%		4%	all
α = .10	all	16%		16%	all

20 observations:

t	0	1	2	3	4	5
P[T=t]	.004	.099	.397	.397	.099	.004
α =.001	12.5%					12.5%
α = .01	all	1%			1%	all
α = .02	all	6%			6%	all
α = .05	all	21%			21%	all
α = .10	all	46%			46%	all

40 observations:

t	0	1	2	3	4	5	6	7	8	9	10
P[T=t]	.000	.000	.011	.078	.239	.344	.239	.078	.011	.000	.000
α =.001	all	all	4.5%						4.5%	all	all
α = .01	all	all	46%						46%	all	all
α = .02	all	all	91%						91%	all	all
α = .05	all	all	all	18%				18%	all	all	all
α = .10	all	all	all	50%				50%	all	all	all

CHAPTER V

EMPIRICAL POWER STUDIES

The worth of a test is determined predominantly by its power, the power function being defined as the probability of rejecting the null hypothesis. The null hypothesis for the test given here is that the underlying distribution is normal. Since all sampling was to be done from distributions simulated on a Spectra 70/45, using Fortran IV subroutines, a program was written to check the random number generator. 2,500 samples of varying sizes were taken from a normal distribution with mean zero and variance one. In order for the random number generator to do a good job, the simulated values and the theoretical values should be quite close. Following is a list of tables comparing the simulated and the theoretical distributions of the test statistic for varying sample sizes. All tables are based on 2,500 observations.

n=8

t	0	1	2
P[T=t]	.167	.666	.167
P[T=t] x 2,500	417.5	1665	417.5
Empirical	428	1665	407

n=12

t	0	1	2	3
P[T=t]	.050	.450	.450	.050
P[T=t] x 2,500	125	1125	1125	125
Empirical	122	1114	1145	119

n=16

t	0	1	2	3	4
P[T=t]	.014	.229	.514	.229	.014
P[T=t] x 2,500	35	572.5	1285	572.5	35
Empirical	26	572	1305	562	35

The above empirical results are remarkably close to the theoretical, indicating that both the random number generator and the program were performing flawlessly. The next step was to simulate densities other than the normal and sample from them.

Programs were written to generate various alternative distributions, sample from them, and apply the proposed test to those samples. The results were tabulated and recorded in the following tables. The entries in the bodies of the tables are the power of the test when sampling from the specified distribution, and the α's are the significance levels. All powers were determined from 1000 observations for each sample size and distribution.

$\chi^2(1)$

Chi-square with one degree of freedom

n \ α	.001	.01	.02	.05	.10
8	.002	.020	.039	.099	.197
12	.005	.046	.093	.232	.463
16	.012	.123	.246	.363	.430
20	.032	.264	.292	.373	.509
40	.401	.567	.747	.815	.872

$\chi^2(2)$

Chi-square with two degrees of freedom

n \ α	.001	.01	.02	.05	.10
8	.002	.016	.032	.080	.160
12	.003	.032	.063	.159	.317
16	.007	.069	.138	.215	.284
20	.015	.126	.151	.226	.351
40	.137	.271	.415	.511	.631

$\chi^2(4)$

Chi-square with four degrees of freedom

n \ α	.001	.01	.02	.05	.10
8	.001	.013	.025	.064	.127
12	.002	.020	.040	.101	.201
16	.004	.038	.076	.128	.195
20	.006	.052	.073	.134	.236
40	.029	.168	.204	.234	.279

$\chi^2(10)$

Chi-square with ten degrees of freedom

n \ α	.001	.01	.02	.05	.10
8	.001	.011	.022	.056	.111
12	.001	.015	.029	.073	.145
16	.002	.021	.042	.079	.137
20	.003	.025	.039	.082	.163
40	.015	.041	.068	.122	.206

Cauchy

n \ α	.001	.01	.02	.05	.10
8	.001	.012	.024	.060	.120
12	.001	.013	.026	.064	.128
16	.001	.013	.026	.056	.112
20	.003	.027	.039	.073	.131
40	.005	.025	.046	.088	.155

U(0,1)

Uniform on interval (0,1)

n \ α	.001	.01	.02	.05	.10
8	.001	.011	.021	.052	.104
12	.001	.012	.025	.062	.124
16	.001	.013	.026	.065	.114
20	.001	.014	.027	.066	.132
40	.001	.014	.028	.063	.119

$$\sum_{i=1}^{2} U_i(0,1)$$

The sum of two U(0,1)'s

n \ α	.001	.01	.02	.05	.10
8	.001	.011	.021	.053	.106
12	.001	.010	.019	.047	.094
16	.001	.010	.019	.046	.102
20	.001	.011	.021	.052	.103
40	.001	.011	.022	.058	.119

Mixed Normal

n \ α	.001	.01	.02	.05	.10
8	.001	.011	.023	.057	.114
12	.002	.018	.034	.086	.172
16	.003	.027	.054	.094	.152
20	.004	.038	.053	.097	.170
40	.010	.034	.061	.111	.192

Values of the mixed normal random variable X were computed as follows. A random variable Y was taken from a normal distribution with mean zero and variance one. If Y were greater than zero then X was taken from a normal with mean five and variance one, and if Y were less than zero then X was taken from a normal with mean minus five and variance one. This distribution was included to check the test's sensitivity to symmetric distributions which are definitely non-normal.

CHAPTER VI

COMPARATIVE ANALYSIS OF THE TESTS

Following is an excerpt from the tables compiled in the article by Shapiro, Wilk, and Chen. [2: 1354-1356] The complete tables will not be given as only parts of them are useful for comparison. In the article the sample sizes used were 10, 15, 20, 35, and 50, while in finding powers for this new test samples of sizes 8, 12, 16, 20, and 40 were used. Many comparisons cannot be made as the sample size for this new test must be some multiple of four. In Table 6.1, n=15 for the nine standard tests and n=16 for the proposed test, while in Table 6.2, n=20 for all tests. All tabled values taken from Shapiro, Wilk, and Chen were arrived at empirically with a 10% level of significance and 200 observations each. In the tables, * represents the test considered in this paper.

TABLE 6.1

10% Powers for n=15

Alternative Distribution	W	$\sqrt{b_1}$	b_2	KS	CM	WCM	D	CS	U	*
U(0,1)	.28	.02	.22	.11	.08	.08	.17	.17	.34	.12
Cauchy	.81	.76	.82	.47	.46	.98	.84	.46	.46	.11
$\chi^2(1)$	.94	.89	.55	.51	.41	.48	.83	.94	.15	.43
$\chi^2(2)$	.82	.71	.38	.30	.24	.26	.39	.43	.18	.28
$\chi^2(4)$	.43	.50	.27	.22	.16	.16	.16	.23	.09	.20
$\chi^2(10)$	.30	.35	.15	.18	.13	.11	.14	.11	.08	.14

TABLE 6.2

10% Powers for n=20

Alternative Distribution	W	$\sqrt{b_1}$	b_2	KS	CM	WCM	D	CS	U	*
U(0,1)	.39	.00	.38	.19	.19	.19	.19	.18	.53	.13
Cauchy	.92	.82	.88	.65	.71	.99	.91	.54	.72	.13
$\chi^2(1)$	.99	.92	.64	.58	.62	.72	.93	.97	.19	.51
$\chi^2(2)$	.89	.82	.44	.37	.33	.41	.56	.43	.12	.35
$\chi^2(4)$	.65	.64	.35	.26	.21	.23	.25	.20	.13	.24
$\chi^2(10)$	.40	.35	.25	.19	.15	.16	.10	.14	.12	.16

Tables 6.1 and 6.2 show that the proposed test is not as sensitive as the other tests on symmetric distributions such as U(0,1) and the Cauchy distribution. When the alternative distribution is asymmetric, as is a Chi-square, then the proposed test is considerably more powerful. When sampling from a $\chi^2(1)$, the proposed test beats only U, but as the degrees of freedom increase the power of the proposed test decreases slower than that of the others, so that when sampling from a $\chi^2(10)$, the proposed test is seen to meet or beat CM, WCM, D, CS, and U.

No comparisons could be made with respect to the sum of two uniforms or the mixed normal since the article quoted [2] did not include these as alternative distributions. The sum of two uniforms is included because this distribution approaches normality rapidly. It is evident from the tables in Chapter V that the test had poor power when sampling from this distribution, but that was to be expected and in fact confirms the rapidity of convergence to the normal. The mixed normal was included since it was symmetric about its mean, though clearly non-normal. Although the test did not show too much sensitivity to the mixed normal, the results do indicate that symmetry need not dictate "poor power".

The test proposed herein is desirable from an experimenter's point of view as it is easy to compute the test statistic. The means are midway between the two observations, while the variance of any pair is one half the squared difference. Finding the medians of such a limited number of means and variances is quite easy and can be done by inspection. Tables for critical regions are readily available in any book that has tables for the hypergeometric distribution.

Another desirable property that the test seems to have is unbiasedness. This means that the power of the test is at least as large as the α-level. That is, the probability of correctly rejecting is at least as large as the probability of incorrectly rejecting. The only exception seen to this appeared when sampling from the sum of two uniforms; however, the powers were so close to the α's that the discrepancy was probably due to sampling error.

Probably the most exciting thing about this test is that it is new and so intuitively satisfying. Basing a test on a complete characterization should result in a good test. The results, however, were disappointing, so an investigation was begun to discern why the powers were not higher. This is the subject of the remaining chapters.

CHAPTER VII

A DISTRIBUTION FREE EXPRESSION FOR $\rho_{\overline{X},S^2}$

The basic question is: If the sample mean and variance are not statistically independent, then how are they related in some specified distribution? Linear relation, being calculable by means of a correlation coefficient, is the obvious start. The usual procedure for calculating the correlation coefficient between the sample mean and variance is to work with the specified density and perform all the integrations and changes of variables necessary. This is a laborious and time-consuming task. What is needed is a formula which can be used for any density, continuous or discrete. Fortunately, it is possible to formulate such an expression for $\rho_{\overline{X},S^2}$ entirely in terms of the central moments of the underlying density. In the pages ahead, that expression is generated and the results applied to eight distributions, four discrete and four continuous. It is hoped that the coming discussion will give some insight into the behavior of the proposed test when sampling from various alternative distributions.

In the proof of the theorem to come the following lemma will be used. Note that all indices of summation run from one to n in this proof, and therefore the limits will be left out.

Lemma.

$$\sum_{i=1}^{n} (X_i-\overline{X})^2 = \sum_{i=1}^{n} (X_i-\mu)^2 - \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} (X_i-\mu)(X_j-\mu)$$

Proof.

$$\Sigma(X_i-\overline{X})^2 = \Sigma[(X_i-\mu) - (\overline{X}-\mu)]^2$$

$$= \Sigma(X_i-\mu)^2 - 2\Sigma(X_i-\mu)(\overline{X}-\mu) + \Sigma(\overline{X}-\mu)^2$$

Now consider

$$\Sigma(X_i-\mu)(\overline{X}-\mu) = \Sigma X_i\overline{X} - \Sigma X_i\mu - \Sigma\mu\overline{X} + \Sigma\mu^2$$

$$= n\overline{X}^2 - \mu n\overline{X} - \mu n\overline{X} + n\mu^2$$

$$= n(\overline{X}-\mu)^2$$

$$= \Sigma(\overline{X}-\mu)^2$$

Therefore,

$$\Sigma(X_i-\overline{X})^2 = \Sigma(X_i-\mu)^2 - 2\Sigma(\overline{X}-\mu)^2 + \Sigma(\overline{X}-\mu)^2$$

$$= \Sigma(X_i-\mu)^2 - \Sigma(\overline{X}-\mu)^2$$

$$= \Sigma(X_i-\mu)^2 - \Sigma(X_i-\mu)(\overline{X}-\mu)$$

$$= \Sigma(X_i-\mu) - (\overline{X}-\mu)\ \Sigma(X_i-\mu)$$

$$= \Sigma(X_i-\mu)^2 - \frac{1}{n}(\Sigma X_j - n\mu)\ \Sigma(X_i-\mu)$$

$$= \Sigma(X_i-\mu)^2 - \frac{1}{n}[\Sigma(X_j-\mu)]\ \Sigma(X_i-\mu)$$

$$= \Sigma(X_i-\mu)^2 - \frac{1}{n}\sum_j\sum_i (X_i-\mu)(X_j-\mu).$$

<u>Theorem</u>. Let $X_1,\ldots,X_n$ be independent and identically distributed random variables. Let $\mu = E(X)$, $\sigma^2 = E[(X-\mu)^2]$, $\mu_3 = E[(X-\mu)^3]$, $\mu_4 = E[(X-\mu)^4]$, $\overline{X} = \Sigma X_i/n$, and $S^2 = \Sigma(X_i-\overline{X})^2/(n-1)$. Then $\rho_{\overline{X},S^2}$, the correlation between $\overline{X}$ and S^2, is given by

$$\rho_{\overline{X},S^2} = \frac{\mu_3}{\sqrt{\sigma^2\left(\mu_4 - \frac{\sigma^4(n-3)}{n-1}\right)}} .$$

Proof.

$$\rho_{\overline{X},S^2} = \frac{\sigma_{\overline{X},S^2}}{\sigma_{\overline{X}}\sigma_{S^2}}$$

$$\sigma_{\overline{X},S^2} = E[\overline{X}S^2] - E[\overline{X}]\ E[S^2]$$

$$= E[\overline{X}S^2] - \mu\sigma^2$$

$$= E\left[\frac{\Sigma X_i}{n} \cdot \frac{1}{n-1}\left(\Sigma(X_i-\overline{X})^2\right)\right] - \mu\sigma^2$$

Application of the preceding lemma yields

$$\sigma_{\overline{X},S^2} = E\left[\frac{\Sigma X_i}{n} \cdot \frac{1}{n-1}\left(\Sigma(X_i-\mu)^2 - \frac{1}{n}\sum_j\sum_i (X_i-\mu)(X_j-\mu)\right)\right] - \mu\sigma^2$$

$$= \frac{1}{n(n-1)}\ E\left[\sum_j\sum_i X_i(X_j-\mu)^2 - \frac{1}{n}\sum_k\sum_j\sum_i X_i(X_j-\mu)(X_k-\mu)\right] - \mu\sigma^2$$

$$= \frac{1}{n(n-1)}\ E\Bigg[\Sigma X_i(X_i-\mu)^2 + \sum_{i\neq j}\sum X_i(X_j-\mu)^2$$

$$- \frac{1}{n}\Bigg\{\Sigma X_i(X_i-\mu)^2 + 2\sum_{i\neq j}\sum X_i(X_i-\mu)(X_j-\mu)$$

$$+ \sum_{i\neq j}\sum X_i(X_j-\mu)^2 + \sum\sum\sum_{\substack{i\neq j\\ j\neq k\\ i\neq k}} X_i(X_j-\mu)(X_k-\mu)\Bigg\}\Bigg] - \mu\sigma^2$$

$$= \frac{1}{n(n-1)} \Bigg[\Sigma E[X_i^3 - 2X_i^2\mu + X_i\mu^2] + \sum_{i\neq j}\sum E[(X_j-\mu)^2]\ E[X_i]$$

$$- \frac{1}{n} \Bigg\{ \Sigma E[X_i^3 - 2X_i^2\mu + X_i\mu^2] + 2 \sum_{i\neq j}\sum E[X_i^2 - X_i\mu]\ E[X_j-\mu]$$

$$+ \sum_{i\neq j}\sum E[X_i]\ E[(X_j-\mu)^2] + \sum_{\substack{i\neq j\\ j\neq k\\ i\neq k}}\sum\sum E[X_i]\ E[X_j-\mu]\ E[X_k-\mu] \Bigg\} \Bigg] - \mu\sigma^2.$$

Note that twice the double sum and the triple sum equal zero since $E[X_i-\mu] = E[X_i] - \mu = \mu-\mu = 0$. Hence we have

$$\sigma_{\overline{X},S^2} = \frac{1}{n(n-1)} \Bigg[\Sigma E[X_i^3 - 2X_i^2\mu + X_i\mu^2] + \sum_{i\neq j}\sum E[X_i]\ E[(X_j-\mu)^2]$$

$$- \frac{1}{n} \Bigg\{ \Sigma E[X_i^3 - 2X_i^2\mu + X_i\mu^2] + \sum_{i\neq j}\sum E[X_i]\ E[(X_j-\mu)^2] \Bigg\} \Bigg] - \mu\sigma^2.$$

Now consider $E[X_i^3 - 2X_i^2\mu + X_i\mu^2]$.

$$\mu_3 = E[(X_i-\mu)^3] = E[X_i^3 - 3X_i^2\mu + 3X_i\mu^2 - \mu^3]$$

$$\mu\sigma^2 = \mu E[(X_i-\mu)^2] = E[X_i^2\mu - 2X_i\mu^2 + \mu^3]$$

Hence $E[X_i^3 - 2X_i^2\mu + X_i\mu^2] = \mu_3 + \mu\sigma^2$.

Using this information the equation becomes

$$\sigma_{\overline{X},S^2} = \frac{1}{n(n-1)}\Bigg[\Sigma(\mu_3 + \mu\sigma^2) + \sum_{i\neq j}\Sigma\ \mu\sigma^2$$

$$- \frac{1}{n}\left\{ \Sigma(\mu_3 + \mu\sigma^2) + \sum_{i\neq j}\Sigma\ \mu\sigma^2 \right\}\Bigg] - \mu\sigma^2$$

$$= \frac{1}{n(n-1)}\Bigg[n(\mu_3 + \mu\sigma^2) + n(n-1)\mu\sigma^2$$

$$- \frac{1}{n}\left(n(\mu_3 + \mu\sigma^2) + n(n-1)\mu\sigma^2 \right)\Bigg] - \mu\sigma^2$$

$$= \frac{1}{n(n-1)} [\mu_3(n-1) + n\mu\sigma^2(n-1)] - \mu\sigma^2$$

$$= \frac{1}{n} [\mu_3 + n\mu\sigma^2] - \mu\sigma^2 = \frac{\mu_3}{n}$$

Therefore,

$$\rho_{\overline{X},S^2} = \frac{\mu_3/n}{\sigma_{\overline{X}}\ \sigma_{S^2}} = \frac{\mu_3/n}{\sigma/\sqrt{n}\ \sigma_{S^2}}$$

$$\sigma^2_{S^2} = E\left[\left\{\frac{\Sigma(X_i-\overline{X})^2}{n-1} - E\left[\frac{\Sigma(X_i-\overline{X})^2}{n-1}\right]\right\}^2\right]$$

$$= E\left[\frac{1}{n-1}\left\{\Sigma(X_i-\mu)^2 - \frac{1}{n}\sum_j\sum_i (X_i-\mu)(X_j-\mu)\right\} - \sigma^2\right]^2$$

$$= E\left[\frac{1}{(n-1)^2}\left\{\Sigma(X_i-\mu)^2\ \Sigma(X_j-\mu)^2 - \frac{2}{n}\Sigma(X_i-\mu)^2 \sum_k\sum_j (X_j-\mu)(X_k-\mu)\right.\right.$$

$$\left.\left. + \frac{1}{n^2}\sum_j\sum_i (X_i-\mu)(X_j-\mu)\sum_k\sum_l (X_k-\mu)(X_l-\mu)\right\}\right] - \sigma^4$$

$$= \frac{1}{(n-1)^2}\quad E\left[\Sigma(X_i-\mu)^4 + \sum\sum_{i\neq j} (X_i-\mu)(X_j-\mu)\right.$$

$$- \frac{2}{n}\sum_{kji}\Sigma\Sigma (X_i-\mu)^2(X_j-\mu)(X_k-\mu)$$

$$\left. + \frac{1}{n^2}\sum_{lkji}\Sigma\Sigma\Sigma (X_i-\mu)(X_j-\mu)(X_k-\mu)(X_i-\mu)\right] - \sigma^4$$

$$= \frac{1}{(n-1)^2} \Bigg[\Sigma E[X_i-\mu]^4 + \sum_{i\neq j}\sum E[X_i-\mu]^2 \; E[X_j-\mu]^2$$

$$- \frac{2}{n} \Bigg\{ \Sigma E[X_i-\mu]^4 + 2 \sum_{i\neq j}\sum E[X_i-\mu]^3 \; E[X_j-\mu]$$

$$+ \sum_{i\neq j}\sum E[X_i-\mu]^2 \; E[X_j-\mu]^2 + \sum_{\substack{i\neq j\\ j\neq k\\ i\neq k}}\sum\sum E[X_i-\mu]^2 E[X_j-\mu] \; E[X_k-\mu] \Bigg\}$$

$$+ \frac{1}{n^2} \Bigg\{ \Sigma E[X_i-\mu]^4 + 4 \sum_{i\neq j}\sum E[X_i-\mu]^3 \; E[X_j-\mu]$$

$$+ 3 \sum_{i\neq j}\sum E[X_i-\mu]^2 \; E[X_j-\mu]^2$$

$$+ \sum_{\substack{i\neq j\\ k\neq l\\ i\neq l}}\sum \sum_{\substack{j\neq k\\ i\neq k\\ j\neq l}}\sum \; E[X_i-\mu] \; E[X_j-\mu] \; E[X_k-\mu] \; E[X_l-\mu] \Bigg\} \Bigg] - \sigma^4$$

In this form, many of the above terms are easily recognized to be equal to zero, and the expression becomes

$$\sigma^2_{S^2} = \frac{1}{(n-1)^2}\left[n\mu_4 + n(n-1)\sigma^4 - \frac{2}{n}\left\{n\mu_4 + n(n-1)\sigma^4\right\}\right.$$

$$\left. + \frac{1}{n^2}\left\{ n\mu_4 + 3n(n-1)\sigma^4\right\} \right] - \sigma^4,$$

which reduces to

$$\sigma^2_{S^2} = \frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)} .$$

Therefore, collecting all this information, we find

$$\rho_{\overline{X},S^2} = \frac{\mu_3/n}{\sqrt{\frac{\sigma^2}{n}\left(\frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}\right)}}$$

$$\rho_{\overline{X},S^2} = \frac{\mu_3}{\sqrt{\sigma^2\left(\mu_4 - \frac{\sigma^4(n-3)}{n-1}\right)}} \qquad (7.1)$$

When $\overline{X}$ and S^2 are independent, which occurs if and only if all the X_i are normal, then $\rho_{\overline{X},S^2} = 0$. Hence, non-zero values of $\rho_{\overline{X},S^2}$ are an indication of non-normality. Unfortunately, however, the converse is not true. A correlation of zero does not imply independence, as correlation measures only the extent of linear relationship among the variables. But because non-zero correlations imply dependence, and hence non-normality, it seems reasonable to use the correlation coefficient between the sample mean and variance as an index of non-normality. In the expression derived for $\rho_{\overline{X},S^2}$, the numerator is μ_3, the third moment about the mean, which is a measure of skewness. For any symmetric distribution μ_3 will equal zero, so for symmetric distributions $\rho_{\overline{X},S^2}$ will be zero, although symmetry does not imply normality. This does not seem too large a concession, though, for $\rho_{\overline{X},S^2}$ may still be used as an index of non-normality for asymmetric distributions.

(7.1) can be written in a form which will lend itself well to the derivations to come.

Consider $\mu_3 = E[X-\mu]^3$

$$= E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3]$$

$$= E[X^3] - 3\mu E[X^2] + 3\mu^2 E[X] - \mu^3$$

$$= E[X^3] - 3\mu E[X^2] + 2\mu^3.$$

In a like manner, it can be shown that

$$\mu_4 = E[X^4] - 4\mu E[X^3] + 6\mu^2 E[X] - 3\mu^4.$$

Therefore, an alternative expression for $\rho_{\overline{X},S^2}$ is given by

$$\rho_{\overline{X},S^2} = \frac{E[X^3] - 3\mu E[X^2] + 2\mu^3}{\sqrt{\sigma^2\left[E[X^4] - 4\mu E[X^3] + 6\mu^2 E[X^2] - 3\mu^4 - \dfrac{\sigma^4(n-3)}{n-1}\right]}}. \qquad (7.2)$$

CHAPTER VIII

$\rho_{\overline{X},S^2}$ FOR VARIOUS ASYMMETRIC DISTRIBUTIONS

In this chapter several asymmetric distributions, both discrete and continuous, are considered and formulas for $\rho_{\overline{X},S^2}$ are derived for each using (7.2). $E[X]$, $E[X^2]$, σ^2, $E[X^3]$, and $E[X^4]$ are calculated for each distribution either by means of the moment-generating-function or the factorial-moment-generating-function, and the results are used to form an expression for $\rho_{\overline{X},S^2}$ in that distribution. Not all the algebra involved is included, as it is lengthy, but for the first distribution considered, the steps involved are outlined and much of the work is shown.

On some of the distributions, comments are made, and on all of them a table of values for $\rho_{\overline{X},S^2}$ is constructed for varying sample sizes and selected values of the parameters involved. These tables are complete enough to give a good idea as to the behavior of $\rho_{\overline{X},S^2}$. However, with the formulas derived, any entry not in the table can be quickly generated.

The density of a Chi-square distribution with k degrees of freedom, $\chi^2(k)$, is

$$f_X(x) = \frac{x^{\frac{k-2}{2}} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} \qquad \text{if } x>0$$

$$= 0 \qquad \text{otherwise,}$$

and its moment-generating-function is given by

$$M_X(t) = (1-2t)^{\frac{-k}{2}}.$$

$E[X^r]$, for any distribution, can be found by differentiating the moment-generating-function r times with respect to t and evaluating the result at t=0. For a $\chi^2(k)$ it can be shown that

$$E[X^r] = \prod_{i=1}^{r} [k + 2(i-1)].$$

Evaluating this we find

$$E[X] = k$$

$$E[X^2] = k(k+2)$$

$$\sigma_X^2 = E[X^2] - E[X]^2 = 2k$$

$$E[X^3] = k(k+2)(k+4)$$

$$E[X^4] = k(k+2)(k+4)(k+6).$$

Substituting this information into (7.2) yields

$$\rho_{\overline{X},S^2} = \frac{k(k+2)(k+4) - 3k(k)(k+2) + 2k^3}{\sqrt{2k\left[k(k+2)(k+4)(k+6)-4k(k+2)(k+4)+6k^2(k)(k+2)-3k^4-\frac{4k^2(n-3)}{n-1}\right]}}$$

$$= \frac{8k}{\sqrt{2k^2\left[12k + 48 - \frac{4k(n-3)}{n-1}\right]}}$$

$$= \frac{4}{\sqrt{6(k+4) - \frac{2k(n-3)}{n-1}}}$$

$$= \frac{4}{\sqrt{\frac{4(nk + 6n - 6)}{n-1}}}$$

$$= \frac{2}{\sqrt{6 + \frac{nk}{n-1}}} .$$

Hence, we have a formula for $\rho_{\overline{X},S^2}$ when sampling from a $\chi^2(k)$. Note that the correlation is a function of both degrees of freedom and sample size as might have been expected.

Table of $\rho_{\overline{X},S^2}$ for a $\chi^2(k)$

k \ n	8	12	16	20	40	∞
1	.748	.751	.752	.753	.755	.756
2	.695	.699	.701	.703	.705	.707
4	.615	.621	.624	.626	.629	.632
10	.479	.486	.490	.492	.496	.500
30	.315	.321	.324	.326	.330	.333
120	.167	.171	.173	.174	.176	.178

The tabled values are seen to decrease as the degrees of freedom increase, which is reasonable when the graph of $\chi^2(k)$ is considered for varying values of k. The graph more closely resembles a normal distribution as k increases although the distribution never becomes symmetric about its mean. $\rho_{\overline{X},S^2}$ is seen to increase with sample size but the increments are negligible.

Consider now the <u>exponential</u> distribution with parameter $\lambda>0$, whose density is

$$f_X(x) = \lambda e^{-x} \quad \text{if } x>0$$

$$= 0 \qquad \text{otherwise.}$$

The moment-generating-function for the exponential is given by

$$M_X(t) = (1 - t/\lambda)^{-1}, \quad t<\lambda.$$

Performing the necessary differentiation, we find

$$E[X] = \frac{1}{\lambda}$$

$$E[X^2] = \frac{2}{\lambda^2}$$

$$\sigma_X{}^2 = \frac{1}{\lambda^2}$$

$$E[X^3] = \frac{6}{\lambda^3}$$

$$E[X^4] = \frac{24}{\lambda^4} .$$

Making use of this information in (7.2) yields

$$\rho_{\overline{X},S^2} = \frac{2}{\sqrt{6 + \frac{2n}{n-1}}} .$$

As might be expected, $\rho_{\overline{X},S^2}$ does not depend on λ since it is only a constant multiplier in the density. Note that this result is the same as that found for a $\chi^2(2)$, as it should be, since they have the same density. Following is the table of values for $\rho_{\overline{X},S^2}$.

n	8	12	16	20	40	∞
$\lambda>0$	.695	.699	.701	.703	.705	.707

The third continuous distribution to be considered is the <u>gamma</u> with parameters $\lambda>0$ and $\theta>0$, $\Gamma(\lambda,\theta)$. Its density takes the form

$$f_X(x) = \frac{\lambda^\theta x^{\theta-1} e^{-\lambda x}}{\Gamma(\theta)} \quad \text{if } x>0$$

$$= 0 \qquad\qquad \text{otherwise.}$$

The moment-generating-function for a $\Gamma(\lambda,\theta)$ is given by

$$M_X(t) = \lambda^{\theta}(\lambda-t)^{-\theta}, \quad \lambda>t.$$

Performing the necessary differentiation we find

$$E[X] = \frac{\theta}{\lambda}$$

$$E[X^2] = \frac{\theta(\theta+1)}{\lambda^2}$$

$$\sigma_X^{\,2} = \frac{\theta}{\lambda^2}$$

$$E[X^3] = \frac{\theta(\theta+1)(\theta+2)}{\lambda^3}$$

$$E[X^4] = \frac{\theta(\theta+1)(\theta+2)(\theta+3)}{\lambda^4} .$$

Substituting into (7.2) we find

$$\rho_{\overline{X},S^2} = \frac{2}{\sqrt{6 + \frac{2\theta n}{n-1}}} .$$

This result is very similar to that found for the exponential. In fact, for $\theta=1$, they are the same. It is easily demonstrated that for $\Gamma(\lambda,1)$, the density reduces to that of an exponential. Again, $\rho_{\overline{X},S^2}$ is found to be independent of the parameter λ, which is reasonable after considering the density.

Table of $\rho_{\overline{X},S^2}$ for a $\Gamma(\lambda,\theta)$

θ \ n	8	12	16	20	40	∞
1/4	.780	.782	.783	.783	.784	.785
1/2	.748	.751	.752	.753	.755	.756
1	.695	.699	.701	.703	.705	.707
2	.615	.621	.624	.626	.629	.632
3	.558	.565	.568	.570	.574	.577
10	.372	.379	.383	.385	.388	.392
100	.131	.134	.135	.136	.138	.139

The final continuous distribution to be considered is the <u>beta</u> with parameters $m>-1$ and $\theta>0$, $\beta(m,\theta)$. The density of a $\beta(m,\theta)$ is given by

$$f_X(x) = \frac{\theta^{m+1} \; x^m e^{-\theta x}}{\Gamma(m+1)} \quad \text{if } x>0$$

$$= 0 \qquad \text{otherwise.}$$

Its moment-generating-function is given by

$$M_X(t) = (1 - \frac{t}{\theta})^{-(m+1)} \quad \text{where } t<\theta.$$

Performing the necessary differentiation yields

$$E[X] = \frac{m+1}{\theta}$$

$$E[X^2] = \frac{(m+1)(m+2)}{\theta^2}$$

$$\sigma_X^{\ 2} = \frac{m+1}{\theta^2}$$

$$E[X^3] = \frac{(m+1)(m+2)(m+3)}{\theta^3}$$

$$E[X^4] = \frac{(m+1)(m+2)(m+3)(m+4)}{\theta^4}.$$

Substitution of this information into (7.2) gives

$$\rho_{\overline{X},S^2} = \frac{2}{\sqrt{6 + \frac{2n(m+1)}{n-1}}}.$$

Again, the similarities between this and the preceding results are striking. These formulas substantiate the facts that $\beta(r-1, \theta) = \Gamma(\lambda,r)$ and also that

$\beta(\frac{r-2}{2}, \frac{1}{2}) = \chi^2(r)$.

Table of $\rho_{\overline{X},S^2}$ for $\beta(m,\theta)$

m \ n	8	12	16	20	40	∞
-.5	.748	.751	.752	.753	.755	.756
0	.695	.699	.701	.703	.705	.707
1	.615	.621	.624	.626	.629	.632
2	.558	.565	.568	.570	.574	.577
5	.450	.457	.461	.463	.467	.471
10	.358	.365	.368	.370	.374	.378
100	.130	.133	.134	.135	.137	.139

The first discrete distribution considered is a binomial with parameters $m=1,2,...$ and $0\leq p\leq 1$, $B(m,p)$. The probability distribution function for a $B(m,p)$ is

$$P_X(x) = \binom{m}{k} p^x q^{m-x}, \quad x = 0,1,2,...,m \text{ where } q=1-p$$

$$= 0 \qquad \text{otherwise.}$$

Its moment-generating-function is given by

$$M_X(t) = (q + pe^t)^m.$$

Performing the necessary differentiation, we find

$$E[X] = mp$$

$$E[X^2] = mp + m(m-1)p^2$$

$$\sigma_X{}^2 = mpq$$

$$E[X^3] = mp + 3m(m-1)p^2 + m(m-1)(m-2)p^3$$

$$E[X^4] = mp + 7m(m-1)p^2 + 6m(m-1)p^3 + m(m-1)(m-2)(m-3)p^4.$$

Making use of this information in (7.2) produces the equation

$$\rho_{\overline{X},S^2} = \frac{q-p}{\sqrt{1 - 6pq + 2mpq\frac{n}{n-1}}} .$$

Tables of $\rho_{\overline{X},S^2}$ for a B(m,p)

m=4

p \ n	8	12	16	20	40	∞
.00	1.00	1.00	1.00	1.00	1.00	1.00
.10	.706	.717	.722	.725	.731	.736
.30	.310	.319	.323	.326	.331	.336
.50	.000	.000	.000	.000	.000	.000
.70	-.310	-.319	-.323	-.326	-.331	-.336
.90	-.706	-.717	-.722	-.725	-.731	-.736
1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00

m=10

p \ n	8	12	16	20	40	∞
.00	1.00	1.00	1.00	1.00	1.00	1.00
.10	.504	.514	.519	.521	.527	.532
.30	.188	.192	.195	.196	.199	.202
.50	.000	.000	.000	.000	.000	.000
.70	-.188	-.192	-.195	-.196	-.199	-.202
.90	-.504	-.514	-.519	-.521	-.527	-.532
1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00

m=30

p \ n	8	12	16	20	40	∞
.00	1.00	1.00	1.00	1.00	1.00	1.00
.10	.311	.317	.321	.323	.327	.330
.30	.106	.109	.110	.111	.112	.114
.50	.000	.000	.000	.000	.000	.000
.70	-.106	-.109	-.110	-.111	-.112	-.114
.90	-.311	-.317	-.321	-.323	-.327	-.330
1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00

The tables show that as m increases $\rho_{\overline{X},S^2}$ approaches zero except at the extremes where p = .00 or p = 1.0. $\rho_{\overline{X},S^2}$ approaches zero even more rapidly as p goes to .5 for any m, which might be expected since a B(m, .5) is symmetric about its mean, and its graph closely approximates the normal N(.5m, .25m).

Closely related to the binomial is the <u>negative binomial</u> with parameters m = 1,2,... and $0\leq p \leq 1$, NB(m,p), whose density is given by

$$P_X(x) = \binom{m+x-1}{x} p^m q^x \quad \text{if } x = 0,1,\ldots \quad \text{where } q=1-p$$

$$= 0 \qquad \text{otherwise.}$$

The moment-generating-function for a NB(m,p) is given by

$$M_X(t) = p^m(1 - qe^t)^{-m}.$$

Performing the necessary differentiation reveals that

$$E[X] = \frac{mq}{p}$$

$$E[X^2] = \frac{mq}{p} + \frac{m(m+1)q}{p^2}$$

$$\sigma_X{}^2 = \frac{mp}{p^2}$$

$$E[X^3] = \frac{mq}{p} + \frac{3m(m+1)q^2}{p^2} + \frac{m(m+1)(m+2)q^3}{p^3}$$

$$E[X^4] = \frac{mq}{p} + \frac{7m(m+1)q^2}{p^2} + \frac{6m(m+1)(m+2)q^3}{p^3} + \frac{m(m+1)(m+2)(m+3)q^4}{p^4}.$$

Upon substituting this information into (7.2) we have

$$\rho_{\overline{X},S^2} = \frac{1+q}{\sqrt{p^2 + 6q + 2mq\,\frac{n}{n-1}}}.$$

Tables of $\rho_{\overline{X},S^2}$ for a NB(m,p)

m=1

p \ n	8	12	16	20	40	∞
.1	.695	.700	.702	.703	.705	.708
.3	.700	.705	.707	.708	.710	.713
.5	.716	.720	.722	.723	.725	.728
.7	.754	.758	.760	.761	.763	.765
.9	.859	.862	.863	.864	.866	.867
1.0	1.00	1.00	1.00	1.00	1.00	1.00

m=2

p \ n	8	12	16	20	40	∞
.1	.616	.622	.625	.626	.630	.633
.3	.621	.627	.630	.632	.635	.638
.5	.638	.644	.646	.648	.651	.655
.7	.679	.685	.688	.690	.693	.696
.9	.805	.810	.812	.813	.815	.818
1.0	1.00	1.00	1.00	1.00	1.00	1.00

m=10

p \ n	8	12	16	20	40	∞
.1	.373	.380	.383	.385	.389	.393
.3	.377	.384	.388	.390	.394	.398
.5	.392	.399	.402	.404	.408	.412
.7	.430	.437	.441	.443	.447	.452
.9	.572	.580	.584	.587	.591	.596
1.0	1.00	1.00	1.00	1.00	1.00	1.00

The next discrete distribution to be considered is the Poisson with parameter $\lambda>0$, $P(\lambda)$. Following is the probability distribution for a $P(\lambda)$:

$$P_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0,1,2,\ldots$$

$$= 0 \qquad \text{otherwise.}$$

The factorial-moment-generating-function for a $P(\lambda)$ is given by

$$\psi_X(t) = e^{\lambda(t-1)}.$$

The factorial-moment-generating-function is related to the moment-generating-function in the following way:

$$\psi_X(e^t) = M_X(t).$$

The necessary differentiation yields

$$E[X] = \lambda$$

$$E[X^2] = \lambda^2 + \lambda$$

$$\sigma_X{}^2 = \lambda$$

$$E[X^3] = \lambda^3 + 3\lambda^2 + \lambda$$

$$E[X^4] = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.$$

Substituting into (7.2) produces the following result:

$$\rho_{\overline{X},S^2} = \frac{1}{\sqrt{1 + 2\lambda\frac{n}{n-1}}}\,.$$

Table of $\rho_{\overline{X},S^2}$ for a $P(\lambda)$

λ \ n	8	12	16	20	40	∞
.25	.798	.804	.808	.809	.813	.816
.5	.683	.692	.696	.698	.703	.707
1	.552	.561	.565	.567	.572	.577
2	.424	.432	.436	.438	.443	.447
3	.357	.364	.368	.370	.374	.378
5	.284	.290	.293	.295	.298	.302
10	.205	.209	.212	.213	.216	.218

The fourth discrete distribution investigated is the _geometric_ with parameter $0 \leq p \leq 1$, $G(p)$, whose density is given by

$$P_X(x) = pq^{x-1} \quad \text{if } x = 1,2,\ldots \quad \text{where } q=1-p$$

$$= 0 \qquad \text{otherwise.}$$

The factorial-moment-generating-function for the geometric is defined to be

$$\psi_X(t) = pt(1-qt)^{-1}.$$

After differentiating and substituting, we find

$$E[X] = 1 + \frac{q}{p}$$

$$E[X^2] = 1 + \frac{3q}{p} + \frac{2q^2}{p^2}$$

$$\sigma_X^2 = \frac{q}{p^2}$$

$$E[X^3] = 1 + \frac{7q}{p} + \frac{12q^2}{p^2} + \frac{6q^3}{p^3}$$

$$E[X^4] = 1 + \frac{15q}{p} + \frac{50q^2}{p^2} + \frac{60q^3}{p^3} + \frac{24q^4}{p^4} .$$

Substituting this information into (7.2), we find

$$\rho_{\overline{X},S^2} = \frac{1+q}{\sqrt{p^2 + 8q + \frac{2q}{n-1}}} .$$

Table of $\rho_{\overline{X},S^2}$ for a G(p)

p \ n	8	12	16	20	40	∞
.00	.695	.699	.701	.703	.705	.707
.10	.695	.700	.702	.703	.705	.708
.30	.700	.705	.707	.708	.710	.713
.50	.716	.720	.722	.723	.725	.728
.70	.754	.758	.759	.761	.763	.765
.90	.859	.862	.863	.864	.866	.867
1.00	1.00	1.00	1.00	1.00	1.00	1.00

The entries in the body of the above table are amazing, for no matter what the value of p, if n is at least eight, then $\rho_{\overline{X},S^2}$ is at least .695, which certainly seems significantly non-zero.

Upon comparing the values of $\rho_{\overline{X},S^2}$ given in the preceding tables to the empirical powers found for a $\chi^2(k)$, which are tabled in Chapter V, it can be seen that the values are not linearly related. There definitely is some degree of relationship, though, since both $\rho_{\overline{X},S^2}$ and the powers increase with larger samples and decrease with added degrees of freedom. This suggests that it might be possible to analyze the behavior of the power function by investigating the relation between $\overline{X}$ and S^2 in a more sophisticated fashion. Regression analysis provides the answer.

If $\overline{X}$ and S^2 are jointly continuous random variables, then the regression function is defined to be the expectation of S^2, given $\overline{X}$, and is calculated by

$$E[S^2|\overline{X}=\overline{x}] = \int_{-\infty}^{\infty} S^2 \frac{f(\overline{X},S^2)}{f(\overline{X})} dS^2.$$

For the above expression to be useful, the joint distribution of $\overline{X}$ and S^2 is needed as well as the marginal density of $\overline{X}$. In order to simplify calculations, only samples of size two will be considered. Letting y_1 and y_2 be a random sample from some specified distribution with density $f(Y)$, the joint distribution becomes $f(y_1,y_2) = f(y_1)f(y_2)$. Using the change of variables $\overline{X} = (y_1+y_2)/2$ and $S^2 = (y_1-y_2)^2/2$, the joint distribution of $\overline{X}$ and S^2 becomes

$$g(\overline{X},S^2) = f(\overline{X},S^2) \frac{1}{|J|},$$

where J is the transformation Jacobian given by $-1/\sqrt{2S^2}$. Using the above method, it is found that the regression of S^2 on $\overline{X}$ for a U(0,1) is given by

$$E[S^2|\overline{X} = \overline{x}] = \begin{cases} \frac{2}{3}\,\overline{x}^2 & \text{if } 0 \le \overline{x} \le \frac{1}{2} \\ \frac{2}{3}\,(1-\overline{x})^2 & \text{if } \frac{1}{2} \le \overline{x} \le 1\,. \end{cases}$$

The graph of this regression equation is definitely not linear, and in fact has a cusp at $\overline{x} = \frac{1}{2}$. (See Figure 8.1.) This fact, along with the poor powers found for the uniform, tends to confirm the fact that if the regression is definitely non-linear, then the power tends to be poor.

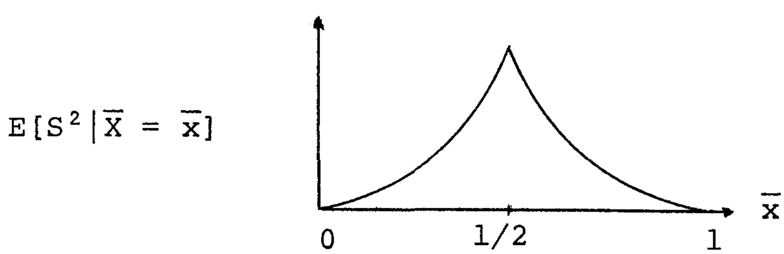

FIGURE 8.1

It is easily seen from Figure 8.1 that no line can be a close approximation for the regression of S^2 on $\overline{X}$ when sampling from a U(0,1).

Performing the same calculations with a $\chi^2(k)$ as the underlying distribution gives rise to the following:

$$E[S^2|\overline{X} = \overline{x}] = \frac{2\overline{x}^2}{k+1} .$$

This function is also non-linear, but much closer to linearity, and the powers are correspondingly much higher than with the U(0,1).

These same techniques could be applied to other densities, but the calculation of the regression function is often non-trivial. The two densities explored have indeed suggested that an approximate indication of the power function can be obtained by considering the regression function.

CHAPTER IX

CONCLUDING REMARKS

In the last few chapters, the behavior of the power function has been investigated in hopes of discerning why the powers of the proposed test were not higher. Consideration of the regression function has provided some insight into this.

It is felt, however, that the proposed test is too weak against symmetric alternatives. Hence, it is appropriate to search for alterations to this test that might yield improvements in power. A possible drawback of the test is the fact that by pairing the observations, the sample size is effectively reduced by a factor of two. A possible alternative would be to form all possible samples of size two from the original sample, thus giving $\binom{n}{2}\bar{X}$ values and $\binom{n}{2}S^2$ values for each sample of size n. This procedure would destroy the independence of $\bar{X}$ and S^2, but a suitable test procedure could still be derived by studying the distribution of the test statistic under normality.

The author has investigated this procedure for a few alternative distributions and has found quite desirable powers. At this time, however, not enough research has been conducted for inclusion in this paper.

Perhaps the test's power would be maximized by pairing the n observations n/2 at a time. This procedure will yield the maximum number of "observations" on which to base the test statistic. For example, if n=10, then 252 $\overline{X}$ values and 252 S^2 values will be generated. Another desirable property of this procedure is that the variances of $\overline{X}$ and S^2 will be greatly reduced.

BIBLIOGRAPHY

1. Kendall, Maurice G. and Stuart, Alan. *The Advanced Theory of Statistics*, Volume 1. New York: Hafner Publishing Company, 1963. 433 pp.

2. Shapiro, S. S., Wilk, M. B., and Chen, H. J. "A Comparative Study of Various Tests for Normality", *Journal of the American Statistical Association*, 63:1343-1372, December, 1968.