CONFIDENCE, PREDICTION, AND TOLERANCE REGIONS FOR THE MULTIVARIATE NORMAL DISTRIBUTION*

VICTOR CHEW
RCA Service Co., Patrick AFB, Florida

Formulas for confidence, prediction, and tolerance regions for the multivariate normal distribution for the various cases of known and unknown mean vector and covariance matrix are assembled for easy reference in this expository paper. Tables are provided for the bivariate case.

1. INTRODUCTION

PROSCHAN [29] gives formulas for confidence, prediction, and tolerance intervals for the univariate normal distribution for all four possible combinations of known and unknown mean and variance. The present paper extends these to the multivariate normal distribution. For expository papers on multivariate normal distributions, see Guenther and Terragno [15] and Gupta [16], the latter with 203 references.

2. NOTATION

The following is a list of symbols that will be used. Matrices and vectors will be indicated by bold-faced letters. Vectors will be column vectors, unless transposed by means of a prime.

p = number of dimensions of the population

$\boldsymbol{\mu} = (\mu_1 \cdots \mu_p)'$ = population mean vector

$\boldsymbol{\Sigma} = (\sigma_{ii'})$ = population covariance matrix of order p

$\mathbf{x} = (x_1 \cdots x_p)'$ = a p-dimensional random vector

$\mathbf{x} = MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ = $\mathbf{x}$ has the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$Q(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ = quadratic form of $\mathbf{x} = MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

n = size of first sample

x_{ij} = value of the ith variable in the jth observation vector $(j=1, \cdots, n; i=1, \cdots, p)$

$\mathbf{x}_j = (x_{1j} \cdots x_{pj})'$ = jth observation vector

$m_i = (x_{i1} + \cdots + x_{in})/n$ = sample mean of ith variable

$\boldsymbol{m} = (m_1 \cdots m_p)' = (\mathbf{x}_1 + \cdots + \mathbf{x}_n)/n$ = sample mean vector

$\boldsymbol{S} = (s_{ii'})$ = sample covariance matrix of order p

$$s_{ii'} = \sum_{j=1}^{n} (x_{ij} - m_i)(x_{i'j} - m_{i'})/(n-1)$$

$\boldsymbol{S}^* \cdot = (s^*_{ii'})$ = sample covariance matrix for known $\boldsymbol{\mu}$

$$s^*_{ii'} = \sum_{j=1}^{n} (x_{ij} - \mu_i)(x_{i'j} - \mu_{i'})/n$$

* This paper is a condensation of the author's report entitled "Confidence, Prediction, and Tolerance Regions," RCA Systems Analysis Technical Memorandum No. 64-9 dated 31 December 1964, prepared under USAF Eastern Test Range, Air Force Systems Command, Contract No. AF 08(606)-5300.

d.f. = degrees of freedom

$\chi^2(\alpha, \nu)$ = upper $(100\alpha)\%$ point of $\chi^2(\nu)$, the Chi-Square distribution with ν d.f., defined by $Pr\{\chi^2(\nu) > \chi^2(\alpha, \nu)\} = \alpha$; and similarly with upper $(100\alpha)\%$ points of other random variables

$\chi'^2(\alpha, \nu, \lambda)$ = upper $(100\alpha)\%$ point of $\chi'^2(\nu, \lambda)$, the non-central Chi-Square distribution with ν d.f. and non-centrality parameter λ

$F(\alpha, \nu_1, \nu_2)$ = upper $(100\alpha)\%$ point of $F(\nu_1, \nu_2)$, the F-distribution with ν_1 and ν_2 d.f.

$t(\alpha, \nu)$ = upper $(100\alpha)\%$ point of $t(\nu)$, the t-distribution with ν d.f.

$z(\alpha)$ = upper $(100\alpha)\%$ point of z, the standard normal distribution

γ = confidence level of probability statement; in the long run, $(100\gamma)\%$ of such statements will be correct

P = proportion of the population to be included in a tolerance region

K = radius of tolerance circle, as a multiple of the larger standard deviation

$h\sigma$ = distance of center of offset tolerance circle from mean of circular normal distribution

r = size of second sample

$\boldsymbol{m}^{(r)}$ = mean vector of second sample

3. CONFIDENCE REGIONS

Based on a random sample $\mathbf{x}_1, \cdots, \mathbf{x}_n$ of n observations from a multivariate normal population, we wish to construct a region R, say, that has $(100\gamma)\%$ probability of enclosing $\boldsymbol{\mu}$. The covariance matrix $\boldsymbol{\Sigma}$ may be known or unknown.

3.1 KNOWN $\boldsymbol{\Sigma}$

If $\mathbf{x} = MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the sample mean vector $\boldsymbol{m} = MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$. The density function of $\boldsymbol{m}$ is proportional to $\exp[-\frac{1}{2}Q(\boldsymbol{m})]$, where

$$Q(\boldsymbol{m}) = (\boldsymbol{m} - \boldsymbol{\mu})'(\boldsymbol{\Sigma}/n)^{-1}(\boldsymbol{m} - \boldsymbol{\mu}) \tag{3.1}$$

has the Chi-Square distribution with p d.f. The hyper-ellipsoid

$$Q(\boldsymbol{m}) = \chi^2(1 - \gamma, p) \tag{3.2}$$

is a $(100\gamma)\%$ confidence region for $\boldsymbol{\mu}$ (more precisely, its boundary).

If $p = 1$, equation (3.2) reduces to $\mu_1 = m_1 \pm \sigma_1\sqrt{\chi^2(1-\gamma, 1)/n}$, the familiar result for the confidence interval for μ, the square root of $\chi^2(1)$ being the standard normal variable.

If $p = 2$, equation (3.2) reduces to

$$n(m_1 - \mu_1 \quad m_2 - \mu_2)\begin{pmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}m_1 - \mu_1 \\ m_2 - \mu_2\end{pmatrix} = \chi^2(1 - \gamma, 2). \tag{3.3}$$

In equation (3.3), every quantity is known numerically except μ_1 and μ_2. The equation is a quadratic in μ_1 and μ_2 and represents an ellipse centered at (m_1, m_2). The orientation and lengths of the axes of the ellipse may be found from standard analytic geometry results.

3.2 UNKNOWN $\boldsymbol{\Sigma}$

From Scheffé [33], page 418, or Anderson [3], page 107, we have that the statistic obtained from (3.1) by replacing $\boldsymbol{\Sigma}$ by S has the Hotelling's T^2-distribution with p and $(n-p)$ d.f. Using the relationship between this and the F-distribution, namely

$$T^2(p, n - p) = \frac{(n - 1)p}{n - p} F(p, n - p), \tag{3.4}$$

the ellipsoid

$$(\boldsymbol{m} - \boldsymbol{\mu})'(S/n)^{-1}(\boldsymbol{m} - \boldsymbol{\mu}) = \frac{(n - 1)p}{n - p} F(1 - \gamma, p, n - p) \tag{3.5}$$

is a $(100\gamma)\%$ confidence region for $\boldsymbol{\mu}$. If $p=1$, equation (3.5) reduces to $\bar{x} \pm ts/\sqrt{n}$, the usual confidence interval for μ, since $F(1, n-1) = t^2(n-1)$.

For Scheffé's or Tukey's simultaneous confidence interval methods, see Scheffé [33]. Roy [31] also discusses this problem.

4. PREDICTION REGIONS

In this section, we discuss the construction of a region R which has $(100\gamma)\%$ probability of containing $\mathbf{x}_{n+1}$, the next single observation or, more generally, the sample mean $\boldsymbol{m}^{(r)}$ of the next r observations, where $\boldsymbol{m}^{(r)} = (\mathbf{x}_{n+1} + \cdots + \mathbf{x}_{n+r})/r$. It is assumed that these next r observations are independent not only of one another but also of the n observations in the first sample.

4.1 KNOWN $\boldsymbol{\mu}$ AND $\boldsymbol{\Sigma}$

Since the population is completely known, the information contained in the first sample $\mathbf{x}_1, \cdots, \mathbf{x}_n$ will not be utilized. From $\boldsymbol{m}^{(r)} = MVN\ (\boldsymbol{\mu}, \boldsymbol{\Sigma}/r)$, the $(100\gamma)\%$ prediction ellipsoid for $\boldsymbol{m}^{(r)}$ is

$$(\boldsymbol{m}^{(r)} - \boldsymbol{\mu})'(\boldsymbol{\Sigma}/r)^{-1}(\boldsymbol{m}^{(r)} - \boldsymbol{\mu}) = \chi^2(1 - \gamma, p), \tag{4.1}$$

In the above equation, only $\boldsymbol{m}^{(r)}$ is unknown. See section 5.1 for spherical and rectangular parallelepipedal prediction regions and for construction of a prediction region that has probability γ of containing at least r_0 out of the next r observations.

In a typical application at the Air Force Eastern Test Range, $r=1$, $p=2$ and elements of $\boldsymbol{\mu}$ are the coordinates (latitude and longitude) of the aimed splash point of a future missile shot. Distribution of actual impact point is assumed to be bivariate normal with $\boldsymbol{\mu}$ as mean vector and $\boldsymbol{\Sigma}$ as covariance matrix, the latter assumed known from past experience. Equation (4.1) is then the equation of an ellipse that has $(100\gamma)\%$ probability of containing the actual splash point of the next missile.

4.2 KNOWN $\boldsymbol{\mu}$ AND UNKNOWN $\boldsymbol{\Sigma}$

Following derivation of (3.5), the $(100\gamma)\%$ prediction ellipsoid for $\boldsymbol{m}^{(r)}$ is

$$(\boldsymbol{m}^{(r)} - \boldsymbol{\mu})'(S^*/r)^{-1}(\boldsymbol{m}^{(r)} - \boldsymbol{\mu}) = \frac{np}{n - p + 1} F(1 - \gamma, p, n - \mathrm{p} + 1). \tag{4.2}$$

4.3 KNOWN $\mathbf{\Sigma}$ AND UNKNOWN $\boldsymbol{\mu}$

In the example of section 4.1, the actual splash point will not be distributed with aim point as the mean value, if systematic errors exist. The unknown mean will now be estimated by $\boldsymbol{m}$, with covariance matrix $\mathbf{\Sigma}/n$.

Based on

$$(\boldsymbol{m}^{(r)} - \boldsymbol{m}) = MVN\left(\mathbf{0}, \frac{n+r}{nr}\mathbf{\Sigma}\right)$$

where $\boldsymbol{m}$ is the mean vector from the first sample, the $(100\gamma)\%$ prediction region for $\boldsymbol{m}^{(r)}$ is

$$\frac{nr}{n+r}(\boldsymbol{m}^{(r)} - \boldsymbol{m})'\mathbf{\Sigma}^{-1}(\boldsymbol{m}^{(r)} - \boldsymbol{m}) = \chi^2(1-\gamma, p). \tag{4.3}$$

4.4 UNKNOWN $\boldsymbol{\mu}$ AND $\mathbf{\Sigma}$

The left-hand side of (4.3) has the T^2-distribution if $\mathbf{\Sigma}$ is replaced by $\boldsymbol{S}$, giving the following $(100\gamma)\%$ prediction region for $\boldsymbol{m}^{(r)}$:

$$\frac{nr}{n+r}(\boldsymbol{m}^{(r)} - \boldsymbol{m})'\boldsymbol{S}^{-1}(\boldsymbol{m}^{(r)} - \boldsymbol{m}) = \frac{(n-1)p}{n-p}F(1-\gamma, p, n-p). \tag{4.4}$$

Recalling that $pF(p, \infty) = \chi^2(p)$, we note that formulas (4.2), (4.3), and (4.4) approach (4.1) as n increases indefinitely.

5. TOLERANCE REGIONS

Two types of a tolerance region R can be distinguished. We can construct R such that the probability is γ that R contains at least $(100P)\%$ of the individuals in the population or such that the average or expected value of the proportion of the population contained in R is exactly $(100P)\%$. We shall henceforth refer to the above tolerance regions as Types 1 and 2 respectively. In terms of repeated sampling, if p_i is the actual proportion contained in R_i, the tolerance region calculated from the ith sample, then $\Pr\{P_i \geqq P\} = \gamma$ or $E(P_i) = P$ according as R is a Type 1 or a Type 2 tolerance region. A $(100P)\%$ Type 2 tolerance region is identical with a $(100P)\%$ prediction region for the next single observation.

5.1 KNOWN $\boldsymbol{\mu}$ AND $\mathbf{\Sigma}$

Because the population parameters are completely known, we can construct a tolerance region R that will, with certainty (i.e., $\gamma = 1$), contain exactly $(100P)\%$ of the population. If $f(\mathbf{x})$ is the joint density function of $\mathbf{x}$, any region R satisfying

$$\int_R f(\mathbf{x})dx = P \tag{5.1}$$

will be a $(100P)\%$ tolerance region. The shape of R may be irregular.

Since R contains exactly $(100P)\%$ of the population, it follows that R is also a $(100P)\%$ prediction region for the next single observation. Also, the probability that R will contain at least r_0 say, out of the next r observations is

$$\sum_{i=r_0}^{i=r} \binom{r}{i} P^i(1-P)^{r-i}.$$

By suitably choosing P, we can make the above probability equal to any preassigned value γ, using tables of the cumulative binomial distribution [19].

5.1.1 *Ellipsoidal regions*

The equation of the $(100P)\%$ tolerance ellipsoid is

$$Q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \chi^2(1 - P, p). \tag{5.2}$$

If $p=1$ and $P=.95$, the above reduces to the usual interval $\mu \pm 1.96\sigma$ that contains 95% of the univariate normal distribution. Equation (5.2) is optimum, in the sense that for any given value of P, the volume of R is a minimum; however, other geometrical shapes for R may be more convenient for other purposes (graphical representation, comparison of several regions, etc.).

Given P, equation (5.2) gives the required tolerance region. Conversely, given that $Q(\mathbf{x})=c^2$ is the equation of a tolerance region R, we can obtain the proportion P of the population contained in it by interpolation in tables of the χ^2-distribution. Harter [18] and Vanderbeck and Cooke [38] tabulate this distribution extensively. The relationship between P and c^2 is

$$\chi^2(1 - P, p) = c^2. \tag{5.3}$$

For $p=2$, integration of the density function of the χ^2-distribution with 2 d.f. reduces equation (5.3) to $P=1-\exp(-c^2/2)$ or $c^2=-2\log_e(1-P)$.

5.1.2 *Spherical regions*

If the distribution of $\mathbf{x}$ is spherical normal (i.e., $\boldsymbol{\Sigma}=\sigma^2 \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix), equation (5.2) is that of a hyper-sphere and the present section is unnecessary.

We can assume, without any loss of generality, that $\boldsymbol{\Sigma}$ is diagonal since otherwise there exists a linear transformation $\boldsymbol{y}=\boldsymbol{A}\mathbf{x}=MVN\ (\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$ such that $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$, the covariance matrix of $\boldsymbol{y}$, is diagonal. One such transformation is given in Wilks [44], page 165. In the bivariate case, this can be achieved by rotating the (x_1, x_2)-axes through an angle θ satisfying

$$\tan(2\theta) = 2\sigma_{12}/(\sigma_1^2 - \sigma_2^2). \tag{5.4}$$

From elementary analytic geometry,

$$\boldsymbol{A} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{5.5}$$

We consider only the bivariate case, for which tables are available. It is also assumed that $\boldsymbol{\Sigma}$ is diagonal; otherwise, the discussion refers to $\boldsymbol{y}=\boldsymbol{A}\mathbf{x}$, where $\boldsymbol{A}$ satisfies equations (5.4) and (5.5). The region R in equation (5.1) is now a circle centered at the mean of the distribution. Let $r=[(x_1-\mu_1)^2+(x_2-\mu_2)^2]^{1/2}$ be the radial distance from the center to a point (x_1, x_2). It is shown in Chew and Boyce [5] that the density function of r is

$$f(r) = (r/\sigma_1\sigma_2) \exp(-ar^2)I_0(br^2), \; r \geqq 0, \tag{5.6}$$

where $a=(\sigma_1^2+\sigma_2^2)/(2\sigma_1\sigma_2)^2$, $b=(\sigma_2^2-\sigma_1^2)/(2\sigma_1\sigma_2)^2$, and $I_0(u)$ is the modified Bessel function of the first kind and zero order. (The above equation reduces to the well-known Rayleigh density function for r if $\sigma_1=\sigma_2$; viz., $f(r)=(r/\sigma^2)$ $\exp[-r^2/(2\sigma^2)]$, since $I_0(0)=1$.)

The circle R will be a $(100P)\%$ tolerance region if its radius is $K\sigma_1$, where

$$P = Pr\{r \leqq K\sigma_1\} = \int_0^{K\sigma_1} f(r)dr. \tag{5.7}$$

Values of K for given values of P and $k=\sigma_2/\sigma_1$ are given in DiDonato and Jarnagin [7] for $k=0(.01)1.00$ and $P=.01(.01).99$; also, for $k=0$, $0.10(.05)1.00$ and $P=.99(.0005).9990(.0001).9999(.00001).99999(.000001).999999$, all to 6 decimals. It is assumed, without losing generality, that $\sigma_1 \geqq \sigma_2$. Smaller tables of $K(P, k)$ are given in Harter [17], for $k=0(.1)1$ and $P=.5$, .75, .90, .95, .975, .99, .995, .9975, and .999, and in Weingarten and DiDonato [42], for $k=.05(.05)1$ and $P=.05(.05).95(.01).99$. An inverse table, giving values of P as a function of K and k, is given in Harter [17] and Esperti [8].

As a numerical example of the loss in efficiency in terms of area when using a circular region, suppose that $\sigma_1=10$ and $\sigma_2=8$. From Harter [17], the radius of the 95% tolerance circle is 22.303, so that its area is $\pi(22.303)^2=497.42\pi$. From elementary analytic geometry, the squares of the semi-major and semi-minor axes of the ellipse $Q(x)=5.99146$ are $\{\sigma_1^2+\sigma_2^2\pm\sqrt{(\sigma_1^2-\sigma_2^2)^2+4\sigma_{12}^2}\}$ $\cdot(5.99146)/2$ or $(24.477)^2$ and $(19.582)^2$ respectively, so that its area is (44.277) $\cdot(19.582)\pi=479.30\pi$. The ratio of the area of the tolerance circle to that of the ellipse is 1.038. Thus, for $P=.95$ and $k=.8$, the loss in efficiency is 3.8%. However, an ellipse requires three parameters for its specification (lengths of major and minor axes and their orientation) while, for fixed center, a circle is completely specified by its radius. This simplifies comparison among a group of tolerance regions. Circles are also more easily represented graphically.

In some applications (see Gilliland [14], for example), we may want the tolerance circle to be off-centered. The region R is then a circle with center (h_1, h_2) relative to (μ_1, μ_2) and radius $K\sigma_1$. Table 1, showing values of K for $P=.95$ as a function of k, h_1, and h_2, has been considerably condensed from DiDonato and Jarnagin [6]. Values of P are tabulated in Germond [12] and Lowe [25] as a function of K, k, h_1, and h_2.

In the particular case of the circular normal distribution ($\sigma_{12}=0$, $\sigma_1=\sigma_2=\sigma$), the proportion P contained in the offset circle depends only on the distance h of the center of the circle from the mean of the distribution and not on the coordinates (h_1, h_2) of the center. The function $P(K, h)$, called the circular coverage function, has been tabulated by Germond [13], Burington and May [4], Table 11.11.1, and Vitalis [39], the last tabulating P for $K/\sigma=0(.01)4.59$ and $h/\sigma=0(.01)3.00$. DiDonato and Jarnagin [7*] tabulates K as a function of P and h.

5.1.3 *Rectangular parallelepipedal regions*

As in the previous section, we assume that $\boldsymbol{\Sigma}$ is diagonal. We have

$$Pr\{\mu_i - z(\tfrac{1}{2} - \tfrac{1}{2}P^{1/p})\sigma_i \leqq x_i \leqq \mu_i + z(\tfrac{1}{2} - \tfrac{1}{2}P^{1/p})\sigma_i\} = P^{1/p}$$
$$i = 1, \cdots, p. \quad (5.8)$$

Because the x's are uncorrelated, the probability that all p statements in equation (5.8) are simultaneously true is P. The points $\mu_i \pm z(\frac{1}{2}-\frac{1}{2}P^{1/p})\sigma_i$, $i=1, \cdots, p$, are thus the vertices of the $(100P)\%$ tolerance rectangular

TABLE 1. RADIUS OF CIRCLE, CENTERED AT (h_1, h_2) RELATIVE TO (μ_1, μ_2), THAT CONTAINS 95% OF THE BIVARIATE NORMAL DISTRIBUTION, WITH $\sigma_{12}=0$, $\sigma_2=1$, AND $\sigma_1 \geq 1$

(Taken from DiDonato and Jarnagin [6], with permission.)

h_1	h_2	σ_1							
		1	2	3	4	5	6	8	10
0	.5	2.5916	4.1849	6.0503	7.9664	9.9007	11.843	15.742	19.649
	1	2.9397	4.4940	6.2781	8.1429	10.043	11.964	15.833	19.723
	2	3.8262	5.3962	7.0506	8.7849	10.583	12.425	16.189	20.010
	3	4.7773	6.3756	8.0040	9.6644	11.371	13.127	16.751	20.475
.25	.5	2.6250	4.1958	6.0563	7.9707	9.9040	11.846	15.744	19.651
	1	2.9646	4.5035	6.2838	8.1471	10.047	11.966	15.835	19.724
	2	3.8407	5.4033	7.0555	8.7887	10.586	12.428	16.190	20.012
	3	4.7873	6.3814	8.0083	9.6678	11.374	13.130	16.753	20.476
.50	.5	2.7195	4.2284	6.0742	7.9834	9.9139	11.854	15.750	19.656
	1	3.0366	4.5319	6.3009	8.1595	10.057	11.974	15.841	19.729
	2	3.8837	5.4247	7.0703	8.8001	10.595	12.436	16.196	20.017
	3	4.8172	6.3988	8.0210	9.6780	11.383	13.137	16.759	20.481
.75	.5	2.8618	4.2821	6.1039	8.0045	9.9305	11.868	15.760	19.664
	1	3.1486	4.5788	6.3293	8.1801	10.073	11.988	15.851	19.737
	2	3.9534	5.4603	7.0949	8.8191	10.611	12.449	16.206	20.025
	3	4.8664	6.4275	8.0423	9.6951	11.397	13.149	16.768	20.489
1	.5	3.0366	4.3563	6.1454	8.0341	9.9536	11.887	15.774	19.675
	1	3.2921	4.6439	6.3688	8.2090	10.096	12.007	15.865	19.748
	2	4.0476	5.5096	7.1292	8.8455	10.632	12.467	16.220	20.036
	3	4.9341	6.4677	8.0721	9.7189	11.417	13.166	16.782	20.499
2	.5	3.8837	4.8298	6.4232	8.2346	10.111	12.017	15.871	19.752
	1	4.0476	5.0674	6.6339	8.4042	10.251	12.135	15.961	19.825
	2	4.6122	5.8370	7.3601	9.0250	10.779	12.590	16.314	20.111
	3	5.3645	6.7358	8.2727	9.8809	11.553	13.283	16.872	20.573
3	.5	4.8172	5.5086	6.8618	8.5574	10.368	12.231	16.031	19.880
	1	4.9341	5.6960	7.0542	8.7200	10.504	12.347	16.120	19.952
	2	5.3645	6.3449	7.7301	9.3167	11.019	12.794	16.469	20.237
	3	5.9871	7.1600	8.5967	10.145	11.776	13.475	17.023	20.696

parallelepiped. If $\boldsymbol{\Sigma}$ is not diagonal, μ_i and σ_i^2 are the mean and variance of y_i, where $\boldsymbol{y}=\boldsymbol{A}\mathbf{x}$ and $\boldsymbol{A\Sigma A'}$ is diagonal, and the above vertices refer to the $\boldsymbol{y}$-axes.

In the bivariate case, the area of the $(100P)\%$ tolerance rectangle is $4\sigma_1\sigma_2 z^2(\frac{1}{2}-\frac{1}{2}\sqrt{P})$. The area of the corresponding ellipse is $\pi\sigma_1\sigma_2\chi^2(1-P, 2)$. The ratios of the former to the latter, for $P=.90$, .95, and .99, are 1.0502, 1.0634, and 1.0885 respectively. To compensate for the loss in efficiency, rectangles are easier to draw; also, separate probability statements can be made about x_1 and x_2 individually.

5.2 UNKNOWN $\boldsymbol{\mu}$ AND $\boldsymbol{\Sigma}$

Wald [40] gives a general parametric method for constructing tolerance regions if sample size n is large, applicable to any given density function. However, the computation is difficult and the adequacy of the method for small samples is unknown.

The following approximate procedure, based on the multivariate normal distribution, is due to John [22]. If $\boldsymbol{m}$ is the sample mean of n observations and $\boldsymbol{S}$ is an unbiased estimate of $\boldsymbol{\Sigma}$, distributed independently of $\boldsymbol{m}$ as a Wishart variable with ν d.f., the tolerance ellipsoid that has approximately $(100\gamma)\%$ probability of containing at least $(100P)\%$ of the population has for its equation

$$(\mathbf{x}-\boldsymbol{m})'\boldsymbol{S}^{-1}(\mathbf{x}-\boldsymbol{m}) = \frac{\chi'^2(1-P,\, p,\, p/n)}{[\chi^2(\gamma,\, \nu p)]/(\nu p)} = H, \text{ say.} \tag{5.9}$$

The approximation is good if $1/n^2$ is negligible. In the usual case where $\boldsymbol{m}$ and $\boldsymbol{S}$ are obtained from the same sample, $\nu=(n-1)$. Equation (5.9) reduces to (5.2) if n and ν are infinite. (It should be noted that equation (2.2) in John [22], which gives the density function of the non-central Chi-Square variable with non-centrality parameter λ, is, in more conventional notation (see, e.g., Wilks [44], page 247), the density with noncentrality parameter (2λ). The non-centrality parameter used in this paper is in accordance with that in Wilks [44].)

To evaluate H, we require the percentage points of the non-central Chi-Square distribution. A recent (unpublished) tabulation is by Haynam and Leone [20]. Small tables are given in Fix [9] and Patnaik [27]. Approximations for the percentage points are given in Abdel-Aty [1], Pearson [28], Roy and Mohamad [30], Sankaran [32], and Tukey [37]. Graphs of $\chi'^2(1-P,\, p,\, p/n)$ are given in Siotani [34] for $p=2$ and 3; $P=.900$, .925, .975, and .990; and $0\leqq(p/n)\leqq .28$.

For the bivariate case, the percentage points are related to the circular coverage function $P(K, h)$, as follows:

$$\alpha = Pr\{\chi'^2(2, \lambda) \leqq \chi'^2(1-\alpha, 2, \lambda)\} = P(\sqrt{\chi'^2(1-\alpha, 2, \lambda}, \sqrt{\lambda}). \tag{5.10}$$

(See Steck [35] and Owen [26].) Using equation (5.10) and interpolating in the table of DiDonato and Jarnagin [7], Appendix E, for values of $\chi'^2(1-P, 2, 2/n)$, Table 2 gives values of H for $p=2$, $\nu=(n-1)$ and P, $\gamma=.90$, .95, and .99.

TABLE 2. VALUES OF H SUCH THAT THE PROBABILITY IS γ (APPROXIMATELY) THAT AT LEAST (100P)% OF THE POPULATION IS CONTAINED IN THE ELLIPSE $(x-m)'S^{-1}(x-m)=H$

γ:	.90			.95			.99		
P: / n	.90	.95	.99	.90	.95	.99	.90	.95	.99
8	9.31	12.08	18.49	11.03	14.32	21.92	15.54	20.17	30.87
9	8.79	11.41	17.48	10.27	13.34	20.43	14.09	18.30	28.02
10	8.39	10.90	16.70	9.70	12.61	19.32	12.99	16.87	25.86
11	8.08	10.50	16.09	9.25	12.02	18.43	12.17	15.81	24.23
12	7.82	10.16	15.58	8.89	11.55	17.72	11.49	14.94	22.91
13	7.60	9.88	15.17	8.59	11.17	17.14	10.97	14.26	21.88
14	7.42	9.65	14.81	8.34	10.84	16.63	10.52	13.68	20.99
15	7.27	9.45	14.51	8.12	10.56	16.21	10.16	13.20	20.26
16	7.13	9.27	14.22	7.92	10.34	15.86	9.83	12.78	19.62
17	7.01	9.12	14.00	7.78	10.12	15.54	9.55	12.42	19.06
18	6.91	8.98	13.80	7.64	9.93	15.25	9.30	12.09	18.57
19	6.81	8.86	13.60	7.51	9.76	14.99	9.08	11.81	18.14
20	6.72	8.74	13.42	7.38	9.60	14.75	8.88	11.54	17.73
22	6.57	8.54	13.12	7.18	9.34	14.36	8.55	11.12	17.08
24	6.45	8.39	12.89	7.03	9.14	14.04	8.27	10.76	16.53
26	6.35	8.26	12.68	6.89	8.96	13.76	8.06	10.48	16.10
28	6.26	8.14	12.50	6.76	8.80	13.51	7.86	10.23	15.72
30	6.18	8.04	12.35	6.66	8.67	13.32	7.70	10.01	15.38
35	6.02	7.84	12.04	6.44	8.38	12.88	7.35	9.56	14.70
40	5.90	7.67	11.79	6.28	8.17	12.56	7.09	9.23	14.18
45	5.80	7.54	11.60	6.15	8.00	12.29	6.90	8.97	13.79
50	5.71	7.43	11.43	6.04	7.86	12.09	6.73	8.76	13.46
60	5.60	7.28	11.19	5.89	7.66	11.77	6.49	8.44	12.98
80	5.43	7.07	10.87	5.68	7.39	11.36	6.17	8.02	12.33
100	5.33	6.93	10.65	5.54	7.20	11.07	5.96	7.76	11.92
200	5.08	6.61	10.16	5.22	6.79	10.43	5.49	7.14	10.98
800	4.83	6.29	9.66	4.90	6.37	9.78	5.02	6.53	10.02
∞	4.61	5.99	9.21	4.61	5.99	9.21	4.61	5.99	9.21

If $(1/n)^2$ is negligible, and for $p=2$, Siotani [34] gives

$$(\mathrm{x} - m)'S^{-1}(\mathrm{x} - m) = [\chi'^2(1 - P, 2, 2/n)]/h \tag{5.11}$$

as the equation of the ellipse that has probability greater than or approximately equal to γ of containing at least $(100P)\%$ of the population, where h satisfies

$$\gamma = 1 - I_{\nu h}(\nu) - \frac{\Gamma[(2\nu + 1)/3]\nu^{(\nu-1)/3}}{\Gamma(\nu)} h^{(\nu-1)/3}[1 - I_{\nu h}\{(2\nu + 1)/3\}]. \tag{5.12}$$

In equation (5.12), $\nu=(n-1)$ and

$$I_\alpha(\nu) = \frac{1}{\Gamma(\nu)} \int_0^\alpha t^{\nu-1}e^{-t}dt \tag{5.13}$$

is the incomplete gamma function [18]. It is rather tedious to solve (5.12) for h.

Fraser and Guttman [10] gives the following as the equation of a Type 2 tolerance ellipsoid:

$$(\mathrm{x} - \boldsymbol{m})'\boldsymbol{S}^{-1}(\mathrm{x} - \boldsymbol{m}) = \frac{\nu p(n+1)F(1-P, p, \nu - p + 1)}{n(\nu - p + 1)}. \tag{5.14}$$

As before, if $\boldsymbol{m}$ and $\boldsymbol{S}$ are estimates from the same sample, $\nu = (n-1)$. Values of the right hand side of (5.14) for $\nu = (n-1)$ and $p = 2$, 3, and 4 are given in the above reference. Equation (5.14) is equivalent to (4.4) if $r = 1$ and $\gamma = P$.

For simultaneous tolerance intervals, see Siotani [34].

5.3 KNOWN $\boldsymbol{\mu}$ AND UNKNOWN $\boldsymbol{\Sigma}$

This may be regarded as a special case of section 5.2 with $n = \infty$ and $\nu = n$. Substituting this into equation (5.9) and replacing $\boldsymbol{m}$ by $\boldsymbol{\mu}$, we have the following equation for the ellipsoid that has probability γ, approximately, of containing at least $(100P)\%$ of the population:

$$(\mathrm{x} - \boldsymbol{\mu})'(\boldsymbol{S}^*)^{-1}(\mathrm{x} - \boldsymbol{\mu}) = \chi^2(1-P, p)/[\chi^2(\gamma, np)/np]. \tag{5.15}$$

Equation (5.15) reduces to (5.2) if $n = \infty$.

For $p = 2$, Siotani [34] replaces the denominator in the right hand side of (5.15) by h, where h is as in (5.12) but with $\nu = n$.

Similarly specializing equation (5.14), the equation of the ellipsoid that will on the average contain exactly $(100P)\%$ of the population is

$$(\mathrm{x} - \boldsymbol{\mu})'(\boldsymbol{S}^*)^{-1}(\mathrm{x} - \boldsymbol{\mu}) = [npF(1-P, p, n-p+1)]/(n-p+1). \tag{5.16}$$

Equation (5.16) is identical with (4.2) if $r = 1$ and $\gamma = P$.

5.4 KNOWN $\boldsymbol{\Sigma}$ AND UNKNOWN $\boldsymbol{\mu}$

Putting $\nu = \infty$ and replacing $\boldsymbol{S}$ by $\boldsymbol{\Sigma}$ in equation (5.14), the equation of the Type 2 tolerance ellipsoid is

$$\begin{aligned}(\mathrm{x} - \boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\mathrm{x} - \boldsymbol{m}) &= \left(\frac{n+1}{n}\right)pF(1-P, p, \infty) \\ &= \left(\frac{n+1}{n}\right)\chi^2(1-P, p).\end{aligned} \tag{5.17}$$

Equations (5.17) and (4.3) are identical if $r = 1$ and $\gamma = P$.

Siotani [34] gives the equation of the Type 1 tolerance ellipsoid as

$$(\mathrm{x} - \boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\mathrm{x} - \boldsymbol{m}) = \chi'^2(1-P, p, \chi^2(1-\gamma, p)/n). \tag{5.18}$$

Equation (5.18) reduces to (5.2) if $n = \infty$. For $p = 2$, equation (5.10) may be used to evaluate the right hand side of (5.18).

6. CONCLUSION

In this paper, it is assumed that the sample has been taken from a single population. Probability regions can also be constructed in the regression situation where the means of the observations are functionally dependent on one or more non-stochastic variables. For example, in the univariate situation, the observation x_i may have expectation $\alpha+\beta t_i$. From the sample (x_i, t_i), $i=1, 2, \cdots, n$, we may wish to construct a probability region for $t=T$. From standard regression analysis, the estimated population mean is a $+bT$, where a and b are the least-squares estimates of α and β respectively. The variance of the estimated mean is σ^2/N, where

$$N^{-1} = \frac{1}{n} + \frac{(T - \bar{t})^2}{\sum (t_i - \bar{t})^2}.$$

N is called the effective number of observations at $t=T$ (Wallis [41]). It is as if we have taken N observations at $t=T$, although we may in fact have taken none at all at $t=T$, and the problem is reduced to the single population case. The problem of probability regions in regression analysis has recently been considered by Weissberg and Beatty [43], Hooper and Zellner [21], Lieberman [23], and Lieberman and Miller [24].

For some recent papers on Bayesian probability regions, see Aitchison [2], Thatcher [36], and Geisser [11].

7. ACKNOWLEDGMENTS

The author wishes to thank the editor and the referees for various comments and suggestions that led to an improvement in the presentation of the paper. Appreciation is also expressed to the U. S. Naval Weapons Laboratory for permission to publish Table 1. Helpful correspondence received from S. John is greatefully acknowledged. Thanks also go to Mrs. Sandra Hubbard for the careful typing of the manuscript.

REFERENCES

[1] Abdel-Aty, S. H., "Approximate Formulae for the Percentage Points and the Probability Integral of the Non-Central Chi-Square Distribution," *Biometrika*, 41 (1954), 538–40.

[2] Aitchison, J., "Bayesian Tolerance Regions," *Journal of the Royal Statistical Society, Series B*, 26 (1964), 161–75.

[3] Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, Inc., 1958.

[4] Burington, R. S. and May, D. C., *Handbook of Probability and Statistics with Tables* New York: McGraw-Hill Book Company, Inc., 1954.

[5] Chew, V. and Boyce, R., "Distribution of Radial Error in the Bivariate Elliptical Normal Distribution," *Technometrics*, 4 (1962), 138–40.

[6] DiDonato, A. R. and Jarnagin, M. P., *Integration of the General Bivariate Gaussian Distribution Over an Offset Ellipse*. Dahlgren, Virginia: U. S. Naval Weapons Laboratory, Report No. 1710, 1960.

[7] DiDonato, A. R. and Jarnagin, M. P., *A Method for Computing the Generalized Circular Error Function and the Circular Coverage Function*. Dahlgren, Virginia: U. S. Naval Weapons Laboratory, Report No. 1768, 1962.

[7*] DiDonato, A. R. and Jarnagin, M. P., "A Method for Computing the Circular Coverage Function," *Mathematics of Computation*, 16 (1962) 347–55.

[8] Esperti, R. V., *Tables of the Elliptical Normal Probability Function*, Santa Barbara, California: Defense Systems Division, General Motors corporation, 1960.

[9] Fix, E., "Tables of Noncentral Chi-Square," *University of California Publications in Statistics*, Volume 1, No. 2 (1949), 15–9.

[10] Fraser, D. A. S. and Guttman, I., "Tolerance Regions," *Annals of Mathematical Statistics*, 27 (1956), 162–79.

[11] Geisser, S., ""Bayesian Estimation in Multivariate Analysis," *Annals of Mathematical Statistics*, 36 (1965), 150–9.

[12] Germond, H. H., *Integration of the Gaussian Distribution Over an Offset Ellipse*. Santa Monica, California: Rand Corporation, Report No. P-94, 1949.

[13] Germond, H. H., *The Circular Coverage Function*. Santa Monica, California: Rand Corporation, Report No. RM-330, (1950).

[14] Gilliland, D. C., "Integral of the Bivariate Normal Distribution Over an Offset Circle," *Journal of the American Statistical Association*, 57 (1962), 758–68.

[15] Guenther, W. C. and Terragno, P. J., "A Review of the Literature on a Class of Coverage Problems," *Annals of Mathematical Statistics*, 35 (1964), 232–60.

[16] Gupta, S. S., "Probability Integrals of Multivariate Normal and Multivariate t and Bibliography," *Annals of Mathematical Statistics*, 34 (1963), 792–838.

[17] Harter, H. Leon, "Circular Error Probabilities," *Journal of the American Statistical Association*, 55 (1960), 723–31.

[18] Harter, H. Leon, *New Tables of the Incomplete Gamma-Function Ratio and of Percentage Points of the Chi-Square Distributions*. Washington, D. C.: U. S. Government Printing Office, 1964.

[19] Harvard Computation Laboratory Staff. *Tables of the Cumulative Binomial Probability Distribution*. Cambridge, Massachusetts: Harvard University Press, 1955.

[20] Haynam, G. E. and Leone, F. C., *Tables of the Non-Central Chi-Square Distribution*. Cleveland, Ohio: Case Institute of Technology, Computing Center and Statistical Laboratory Report No. 1060, 1962.

[21] Hooper, J. and Zellner, A., "The Error of Forecast for Multivariate Regression Models," *Econometrica*, 29 (1961), 544–55.

[22] John, S., "A Tolerance Region for Multivariate Normal Distributions," *Sankhyā*, 25 (1963) 363–8.

[23] Lieberman, G. J., "Prediction Regions for Several Predictions from a Single Regression Line," *Technometrics*, 3 (1961), 21–7.

[24] Lieberman, G. J. and Miller, R. G., Jr., "Simultaneous Tolerance Intervals in Regression," *Biometrika*, 50 (1963), 155–68.

[25] Lowe, J. R., "A Table of the Integral of the Bivariate Normal Distribution Over an Offset Circle," *Journal of the Royal Statistical Society, Series B*, 22 (1960), 177–87.

[26] Owen, D. B., *Handbook of Statistical Tables*. Reading, Massachusetts: Addison-Wesley Publishing Co., Inc., 1962.

[27] Patnaik, P. B., "The Non-Central Chi-Square and F-Distributions and Their Applications," *Biometrika*, 36 (1949), 202–32.

[28] Pearson, E. S., "Note on an Approximation to the Distribution of Non-Central Chi-Square," *Biometrika*, 46 (1959), 364.

[29] Proschan, F., "Confidence and Tolerance Intervals for the Normal Distribution," *Journal of the American Statistical Association*, 48 (1953), 550–64.

[30] Roy, J. and Mohamad, J., "An Approximation to the Non-Central Chi-Square Distribution," *Sankhyā*, 26 (1964), 81–4.

[31] Roy, S. N., *Some Aspects of Multivariate Analysis*. New York: John Wiley and Sons, Inc., 1957.

[32] Sankaran, M., "On the Non-Central Chi-Square Distribution," *Biometrika*, 46 (1959), 235–7.

[33] Scheffé, H., *The Analysis of Variance*. New York: John Wiley and Sons, Inc., 1959.

[34] Siotani, M., "Tolerance Regions for a Multivariate Normal Population," *Annals of the Institute of Statistical Mathematics*, 16 (1964), 135–53.

[35] Steck, G. P., *Uses of the Table of the Circular Coverage Function*. Washington, D. C.: Office of Technical Services, Dept. of Commerce, Sandia Corporation Technical Memorandum No. SCTM-214-58(51), 1958.

[36] Thatcher, A. R., "Relationships Between Bayesian and Confidence Limits for-Predictions," *Journal of the Royal Statistical Society*, Series B, 26 (1964), 176–210.

[37] Tukey, J. W., "Approximation to the Upper 5% Points of Fisher's B Distribution and Non-Central χ^2," *Biometrika*, 44 (1957), 528–30.

[38] Vanderbeck, J. P. and Cooke, J. R., *Extended Table of Percentage Points of the Chi-Square Distribution*. China Lake, California: U. S. Naval Ordnance Test Station NAVWEPS Report 7770, NOTS TP2749, 1961.

[39] Vitalis, J. A., *Table of Circular Normal Probabilities*. Buffalo, New York: Operations Analysis Group, Dynamics Section, Bell Aircraft Corporation, 1956. (Astia Document No. AD-139-515.)

[40] Wald, A., "Setting of Tolerance Limits When the Sample Size is Large," *Annals of Mathematical Statistics*, 13 (1942), 389–99.

[41] Wallis, W. A., "Tolerance Intervals for Linear Regession." *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Ed., J. Neyman), 43–51. Berkeley: University of California Press, 1951.

[42] Weingarten, H. and DiDonato, A. R., "A Table of Generalized Circular Error," *Mathematics of Computation*, 15 (1961), 169–73.

[43] Weissberg, A. and Beatty, G. H. "Tables of Tolerance-Limit Factors for Normal Distributions," *Technometrics*, 2 (1960) 483–500.

[44] Wilks, S. S. *Mathematical Statistics*. New York: John Wiley and Sons, Inc., 1962.