Complex Random Variables

Leigh J. Halliwell, FCAS, MAAA

Abstract: Rarely have casualty actuaries needed, much less wanted, to work with complex numbers. One readily could wisecrack about imaginary dollars and creative accounting. However, complex numbers are well established in mathematics; they even provided the impetus for abstract algebra. Moreover, they are essential in several scientific fields, most notably in electromagnetism and quantum mechanics, the two fields to which most of the sparse material about complex random variables is tied. This paper will introduce complex random variables to an actuarial audience, arguing that complex random variables will eventually prove useful in the field of actuarial science. First, it will describe the two ways in which statistical work with complex numbers differs from that with real numbers, viz., in transjugation versus transposition and in rank versus dimension. Next, it will introduce the mean and the variance of the complex random vector, and derive the distribution function of the standard complex normal random vector. Then it will derive the general distribution of the complex normal multivariate and discuss the behavior and moments of complex lognormal variables, a limiting case of which is the unit-circle random variable $W = e^{i\Theta}$ for real Θ uniformly distributed. Finally, it will suggest several foreseeable actuarial applications of the preceding theory, especially its application to linear statistical modeling. Though the paper will be algebraically intense, it will require little knowledge of complex-function theory. But some of that theory, viz., Cauchy's theorem and analytic continuation, will arise in an appendix on the complex moment generating function of a normal random multivariate.

Keywords: Complex numbers, matrices, and random vectors; augmented variance; lognormal and unit-circle distributions; determinism; Cauchy-Riemann; analytic continuation

1. INTRODUCTION

Even though their education has touched on algebra and calculus with complex numbers, most casualty actuaries would be hard-pressed to cite an actuarial use for numbers of the form $x + iy$. Their use in the discrete Fourier transformation (Klugman [1998], §4.7.1) is notable; however, many would view this as a trick or convenience, rather than as indicating any further usefulness. In this paper we will develop a probability theory for complex random variables and vectors, arguing that such a theory will eventually find actuarial uses. The development, lengthy and sometimes arduous, will take the following steps. Sections 2-4 will base complex matrices in certain real-valued matrices called "double-real." This serves the aim of our presentation, namely, to analogize from real-valued random variables and vectors to complex ones. Transposition and dimension in the real-valued

realm become transjugation and rank in the complex. These differences figure into the standard quadratic form of Section 5, where also the distribution of the standard complex normal random vector is derived. Section 6 will elaborate on the variance of a complex random vector, as well as introduce "augmented variance," i.e., the variance of dyad whose second part is the complex conjugate of the first. Section 7 derives of the formula for the distribution of the general complex normal multivariate. Of special interest to many casualty actuaries should be the treatment of the complex lognormal random vector in Section 8, an intuition into whose behavior Section 9 provides on a univariate or scalar level. Even further simplification in the next two sections leads to the unit-circle random variable, which is the only random variable with widespread deterministic effects. In Section 12 we adapt the linear statistical model to complex multivariates. Finally, Section 13 lists foreseeable applications of complex random variables. However, we believe their greatest benefit resides not in their concrete applications, but rather in their fostering abstractions of thought and imagination. Three appendices delve into mathematical issues too complicated for the body of paper. Those who work on an advanced level with lognormal random variables should read Appendix A ("Real-Valued Lognormal Random Vectors"), regardless of their interest in complex random variables.

2. INVERTING COMPLEX MATRICES

Let $m \times n$ complex matrix Z be composed of real and imaginary parts X and Y, i.e., $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$. Of course, X and Y also must be $m \times n$. Since only square matrices have inverses, our purpose here requires that $m = n$. Complex matrix $\mathbf{W} = \mathbf{A} + i\mathbf{B}$ is an inverse of Z if and only if $\mathbf{ZW} = \mathbf{WZ} = \mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Because such an inverse must be unique, we may say that $\mathbf{Z}^{-1} = \mathbf{W}$. Under what conditions does W exist?

First, define the conjugate of Z as $\overline{Z} = X - iY$. Since the conjugate of a product equals the product of the conjugates,[1] if Z is non-singular, then $\overline{Z}\overline{Z^{-1}} = \overline{ZZ^{-1}} = \overline{I_n} = I_n$. Similarly, $\overline{Z^{-1}}\overline{Z} = I_n$.

Therefore, $\overline{Z}$ too is non-singular, and $\overline{Z}^{-1} = \overline{Z^{-1}}$. Moreover, if Z is non-singular, so too are $i^n Z$ and $i^n \overline{Z}$. Therefore, the invertibility of $X + iY$, $-Y + iX$, $-X - iY$, $Y - iX$, $X - iY$, $Y + iX$, $-X + iY$, and $-Y - iX$ is true for all eight or true for none. Invertibility is no respecter of the real and imaginary parts.

Now if the inverse of Z is $W = A + iB$, then $I_n = (X + iY)(A + iB) = (A + iB)(X + iY)$. Expanding the first equality, we have:

$$\begin{aligned}
I_n &= (X + iY)(A + iB) \\
&= XA + iYA + iXB + i^2 YB \\
&= XA + iYA + iXB - YB \\
&= (XA - YB) + i(YA + XB)
\end{aligned}$$

Therefore, $ZW = I_n$ if and only if $XA - YB = I_n$ and $YA + XB = 0_{n \times n}$. We may combine the last two equations into the partitioned-matrix form:

$$\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

Since $YA + XB = 0$ if and only if $-YA - XB = 0$, another form just as valid is:

$$\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}\begin{bmatrix} -B \\ A \end{bmatrix} = \begin{bmatrix} 0 \\ I_n \end{bmatrix}$$

We may combine these two forms into the balanced form:

$$\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}\begin{bmatrix} A & -B \\ B & A \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix} = I_{2n}$$

---

[1] If Z and W are conformable to multiplication:
$$\overline{ZW} = \overline{(X + iY)(A + iB)} = \overline{XA - YB + i(XB + YA)} = XA - YB - i(XB + YA) = (X - iY)(A - iB) = \overline{Z}\overline{W}$$

Therefore, $\mathbf{ZW} = \mathbf{I}_n$ if and only if $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}\begin{bmatrix} A & -B \\ B & A \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix}$. By a similar expansion of the

last equality above, $\mathbf{WZ} = \mathbf{I}_n$ if and only if $\begin{bmatrix} A & -B \\ B & A \end{bmatrix}\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix}$. Hence, we conclude

that the $n{\times}n$ complex matrix $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ is non-singular, or has an inverse, if and only if the $2n{\times}2n$

real-valued matrix $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ is non-singular. Moreover, if $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}^{-1}$ exists, it will have the form

$\begin{bmatrix} A & -B \\ B & A \end{bmatrix}$ and $\mathbf{Z}^{-1}$ will equal $\mathbf{A} + i\mathbf{B}$.

3. COMPLEX MATRICES AS DOUBLE-REAL MATRICES

That the problem of inverting an $n{\times}n$ complex matrix resolves into the problem of inverting a

$2n{\times}2n$ real-valued matrix suggests that with complex numbers one somehow gets "two for the price

of one." It even hints of a relation between the general $m{\times}n$ complex matrix $\mathbf{Z} = \mathbf{X} + i\mathbf{Y}$ and the

$2m{\times}2n$ complex matrix $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$. If X and Y are $m{\times}n$ real matrices, we will call $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ a

double-real matrix. The matrix is double in two senses; first, in that it involves two (same-sized and

real-valued) matrices X and Y, and second, in that its right half is redundant, or reproducible from

its left.

Returning to the hint above, we easily see an addition analogy:

$$X_1 + iY_1 + X_2 + iY_2 \Leftrightarrow \begin{bmatrix} X_1 & -Y_1 \\ Y_1 & X_1 \end{bmatrix} + \begin{bmatrix} X_2 & -Y_2 \\ Y_2 & X_2 \end{bmatrix}$$

And if Z_1 is $m \times n$ and Z_2 is $n \times p$, so that the matrices are conformable to multiplication, then

$$Z_1 Z_2 = (X_1 + iY_1)(X_2 + iY_2) = (X_1 X_2 - Y_1 Y_2) + i(X_1 Y_2 + Y_1 X_2).$$ This is analogous with the double-real multiplication:

$$\begin{bmatrix} X_1 & -Y_1 \\ Y_1 & X_1 \end{bmatrix} \begin{bmatrix} X_2 & -Y_2 \\ Y_2 & X_2 \end{bmatrix} = \begin{bmatrix} X_1 X_2 - Y_1 Y_2 & -X_1 Y_2 - Y_1 X_2 \\ X_1 Y_2 + Y_1 X_2 & X_1 X_2 - Y_1 Y_2 \end{bmatrix}$$

Rather trivial is the analogy between the $m \times n$ complex zero matrix and the $2m \times 2n$ double-real zero

matrix $\begin{bmatrix} 0_{m \times n} & -0 \\ 0 & 0 \end{bmatrix}$, as well as that between the $n \times n$ complex identity matrix and the $2n \times 2n$ double-

real identity matrix $\begin{bmatrix} I_n & -0 \\ 0 & I_n \end{bmatrix}$.

The general $2m \times 2n$ double-real matrix may itself be decomposed into quasi-real and quasi-imaginary

parts: $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} + \begin{bmatrix} 0 & -Y \\ Y & 0 \end{bmatrix}$. And in the case of square matrices $(m = n)$ this extends

to the form $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} + \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix} \begin{bmatrix} Y & 0 \\ 0 & Y \end{bmatrix}$, wherein the double-real matrix

$\begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix}$ is analogous with the imaginary unit, inasmuch as:

$$\begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix} \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix} = \begin{bmatrix} -I_n & 0 \\ 0 & -I_n \end{bmatrix} = (-1) \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix}$$

Finally, one of the most important theorems of linear algebra is that every $m \times n$ complex matrix $Z = X + iY$ may be reduced by invertible transformations to "canonical form" (Healy [1986], 32-34). In symbols, for every Z there exist non-singular matrices U and V such that:

$$U_{m \times m} Z_{m \times n} V_{n \times n} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$$

The *m×n* real matrix on the right side of the equation consists entirely of zeroes except for *r* instances of one along its main diagonal. Since invertible matrix operations can reposition the ones, it is further stipulated that the ones appear as a block in the upper-left corner. Although many reductions of Z to canonical form exist, the canonical forms themselves must all contain the same number of ones, *r*, which is defined as the rank of Z. Providing the matrices with real and complex parts, we have:

$$\begin{aligned}
U_{m \times m} Z_{m \times n} V_{n \times n} &= (P + iQ)(X + iY)(R + iS) \\
&= (PXR - QYR - PYS - QXS) + i(PXS - QYS + PYR + QXR) \\
&= A + iB \\
&= \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} + i0_{m \times n}
\end{aligned}$$

The double-real analogue to this is:

$$\begin{bmatrix} P & -Q \\ Q & P \end{bmatrix} \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} R & -S \\ S & R \end{bmatrix} = \begin{bmatrix} A & -B \\ B & A \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} & 0_{m \times n} \\ 0_{m \times n} & \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix}$$

As shown in the previous section, $\begin{bmatrix} P & -Q \\ Q & P \end{bmatrix}$ is non-singular, or invertible, if and only if $U = P + iQ$ is non-singular; the same is true for $\begin{bmatrix} R & -S \\ S & R \end{bmatrix}$. Therefore, the rank of the double-real analogue of a complex matrix is twice the rank of the complex matrix. Moreover, the 2*r* instances of one correspond to *r* quasi-real and *r* quasi-imaginary instances. It is not possible for the contribution to the rank of a matrix to be real without its being imaginary, and *vice versa*.

To conclude this section, there are extensive analogies between complex and double-real matrices, analogies so extensive that one who lacked either the confidence or the software to work with complex numbers could probably do a work-around with double-real matrices.[2]

4. COMPLEX MATRICES AND VARIANCE

$\Sigma = Var[\mathbf{x}] = E\left[(\mathbf{x} - \mu)(\mathbf{x} - \mu)' \right]$ is a real-valued $n \times n$ matrix, whose jk^{th} element is the covariance of

the jth element of x with the kth element. Since the covariance of two real-valued random variables is symmetric, Σ must be a symmetric matrix. But a realistic Σ must have one other property, viz., non-negative definiteness (NND). This means that for every real-valued $n \times 1$ vector ξ, $\xi' \Sigma \xi \geq 0$.[3] This

must be true, because $\xi' \Sigma \xi$ is the variance of the real-valued random variable $\xi' \mathbf{x}$:

$$Var[\xi'\mathbf{x}] = E\left[(\xi'\mathbf{x} - \xi'\mu)(\xi'\mathbf{x} - \xi'\mu)' \right] = E\left[\xi'(\mathbf{x} - \mu)(\mathbf{x} - \mu)' \xi \right] = \xi'E\left[(\mathbf{x} - \mu)(\mathbf{x} - \mu)' \right]\xi = \xi'\Sigma\xi$$

Although variances of real-valued random variables may be zero, they must not be negative. Now if $\xi' \Sigma \xi > 0$ for all $\xi \neq 0_{n \times 1}$, the variance Σ is said to be positive-definite (PD). Every invertible NND matrix must be PD. Moreover, every NND matrix may be expressed as the product of some real matrix and its transpose, the most common method for doing this being the Cholesky

---

[2] The representation of the complex scalar $z = x + iy$ as the real 2×2 matrix $\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$ is a common theme in modern algebra (e.g., section 7.2 of the Wikipedia article "Complex number"). We have merely extended the representation to complex matrices. Our representation is even more meaningful when expressed in the Kronecker-product form $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} + \begin{bmatrix} 0 & -Y \\ Y & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes X + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes Y.$ Due to certain properties of the Kronecker product (cf. Judge [1988], Appendix A.15), all the analogies of this section would hold even in the commuted form $X \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + Y \otimes \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$ In practical terms this means that it matters not whether the form is 2×2 of $m \times n$ or $m \times n$ of 2×2.

[3] More accurately, $\xi' \Sigma \xi \geq [0]$, since the quadratic form $\xi' \Sigma \xi$ is a 1×1 matrix. The relevant point is that 1×1 real-valued matrices are as orderable as their real-valued elements are. Appendices A.13 and A.14 of Judge [1988] provide introductions to quadratic forms and definiteness that are sufficient to prove the theorems used herein.

decomposition (Healy [1986], §7.2). Accordingly, if Σ is NND, then $A'\Sigma A \geq 0$ for any conformable real-valued matrix A. Finally, if Σ is PD and real-valued $n \times r$ matrix A is of full column rank, i.e., $rank(A_{n \times r}) = r$, then the $r \underline{\times} r$ matrix $A'\Sigma A$ is PD.

In the remainder of this section we will show how the analogy between $X + iY$ and $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ leads to a proper definition of the variance of a complex random vector. We start by considering $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ as a real-valued variance matrix. In order to be so, first it must be symmetric:

$$\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}' = \begin{bmatrix} X' & Y' \\ -Y' & X' \end{bmatrix} = \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$$

Hence, $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ is symmetric if and only if $X' = X$ and $Y' = -Y$. In words, X is symmetric and Y is skew-symmetric. Clearly, the main diagonal of a skew-symmetric matrix must be zero. But of greater significance, if a and b are real-valued $n \times 1$ vectors:

$$a'Yb = (a'Yb)_{1 \times 1} = (a'Yb)' = b'Y'a = b'(-Y)a = -b'Ya$$

Consequently, if $b = a$:

$$a'Ya = \frac{a'Ya + a'Ya}{2} = \frac{a'Ya + (-a'Ya)}{2} = 0_{1 \times 1}$$

Next, considering the specifications on X, Y, a, and b, we evaluate the 2×2 quadratic form:

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a' & b' \\ -b' & a' \end{bmatrix} \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

$$= \begin{bmatrix} a'X + b'Y & b'X - a'Y \\ -b'X + a'Y & a'X + b'Y \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

$$= \begin{bmatrix} a'Xa + b'Ya - a'Yb + b'Xb & -a'Xb - b'Yb + b'Xa - a'Ya \\ -b'Xa + a'Ya + a'Xb + b'Yb & a'Xa + b'Ya - a'Yb + b'Xb \end{bmatrix}$$

$$= \begin{bmatrix} a'Xa + b'Ya - a'Yb + b'Xb & -a'Xb - 0 + b'Xa - 0 \\ -b'Xa + 0 + a'Xb + 0 & a'Xa + b'Ya - a'Yb + b'Xb \end{bmatrix}$$

$$= \begin{bmatrix} a'Xa + b'Ya - a'Yb + b'Xb & -a'Xb + b'Xa \\ -b'Xa + a'Xb & a'Xa + b'Ya - a'Yb + b'Xb \end{bmatrix}$$

$$= \begin{bmatrix} a'Xa + b'Ya - a'Yb + b'Xb & 0 \\ 0 & a'Xa + b'Ya - a'Yb + b'Xb \end{bmatrix}$$

$$= \begin{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} & 0 \\ 0 & \begin{bmatrix} a \\ b \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \end{bmatrix}$$

Therefore, $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ is PD [or NND] if and only if $\begin{bmatrix} a \\ b \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ is PD

[or NND].

Now the double-real $2n \times 2$ matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ is analogous with the $n \times 1$ complex vector $a + ib$. Its

transpose $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}' = \begin{bmatrix} a' & b' \\ -b' & a' \end{bmatrix}$ is analogous with the $1 \times n$ complex vector $a' - ib'$. Moreover,

$a' - ib' = (a - ib)' = \overline{a + ib}' = \overline{(a + ib)}' = (a + ib)^*$, where '*' is the combined operation of

transposition and conjugation (order irrelevant).[4] And $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ is analogous with the $n{\times}n$

complex matrix $X + iY$. Accordingly, the complex analogue of the double-real quadratic form

$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ is $(a + ib)^* (X + iY)(a + ib)$. Moreover, since $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ is

symmetric, $(X + iY)^* = X' - iY' = X - i(-Y) = X + iY$. A matrix equal to its transposed conjugate

is said to be Hermetian: matrix Γ is Hermetian if and only if $\Gamma^* = \Gamma$. Therefore, $\Gamma_{n{\times}n} = X + iY$ is

the variance matrix of some complex random variable $\mathbf{z}_{n{\times}1} = \mathbf{x} + i\mathbf{y}$ if and only if Γ is Hermetian

and $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ is non-negative-definite.[5]

Because $(a + ib)^* (X + iY)(a + ib) = \begin{bmatrix} a \\ b \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + i \cdot 0 = \begin{bmatrix} a \\ b \end{bmatrix}' \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$, the definiteness

of $\Gamma = X + iY$ is the same as the definiteness of $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$. Therefore, a matrix qualifies as the

variance matrix of some complex random vector if and only if it is Hermetian and NND. Just as the

variance matrix of a real-valued random vector factors as $\Sigma = A'A_{n{\times}n}$ for some real-valued A, so too

the variance matrix of a complex random vector factors as $\Gamma = A^* A_{n{\times}n}$ for some complex A.

Likewise, every invertible Hermetian NND matrix must be PD. Due to the skew symmetry of their

---

[4] The transposed conjugate is sometimes called the "transjugate," which in linear algebra is commonly symbolized with the asterisk. Physicists prefer the "dagger" notation A†, though physicist Hermann Weyl [1950, p. 17] called it the "Hermetian conjugate" and symbolized it as $\tilde{A}$.

[5] It is superfluous to add 'symmetric' here. For $\Gamma = X + iY$ is Hermetian if and only if $\begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$ is symmetric.

complex parts, the main diagonals of Hermetian matrices must be real-valued. If the matrices are NND [or PD], all the elements of their main diagonals must be non-negative [or positive].

Let Γ represent the variance of the complex random vector $\mathbf{z}$. Its jk^{th} element represents the covariance of the jth element of $\mathbf{z}$ with the kth element. Since Γ is Hermetian, $\gamma_{kj} = [\Gamma]_{kj} = [\Gamma^*]_{kj} = [\overline{\Gamma}']_{kj} = [\overline{\Gamma}]_{jk} = [\Gamma]_{jk} = \overline{\gamma}_{jk}$. Because of this, it is fitting and natural to define the variance of a complex random vector as:

$$\Gamma = Var[\mathbf{z}] = E\left[(\mathbf{z} - \mu)\overline{(\mathbf{z} - \mu)}' \right] = E\left[(\mathbf{z} - \mu)(\mathbf{z} - \mu)^* \right]$$

The complex formula is like the real formula except that the second factor in the expectation is transjugated, not simply transposed. This renders Γ Hermetian, since:

$$\Gamma^* = E\left[(\mathbf{z} - \mu)(\mathbf{z} - \mu)^* \right]^* = E\left[\{ (\mathbf{z} - \mu)(\mathbf{z} - \mu)^* \}^* \right] = E\left[(\mathbf{z} - \mu)(\mathbf{z} - \mu)^* \right] = \Gamma$$

It also renders Γ NND. For since $(\mathbf{z} - \mu)(\mathbf{z} - \mu)^*$ is NND, its expectation over the probability distribution of $\mathbf{z}$ must also be so. Usually Γ is PD, in which case Γ^{-1} exists.

5. THE EXPECTATION OF THE STANDARD QUADRATIC FORM

The most common quadratic form in $\mathbf{z}_{n\times 1}$ involves the variance of the complex random variable, viz., $(\mathbf{z} - \mu)^* \Gamma^{-1} (\mathbf{z} - \mu)$, where $\Gamma = Var[\mathbf{z}]$. The expectation of this quadratic form equals n, the rank of the variance. The following proof uses the trace function. The trace of a matrix is the sum of its main-diagonal elements, and if A and B are conformable $tr(AB) = tr(BA)$. Moreover, the trace of the expectation equals the expectation of the trace. Consequently:

$$E\left[(\mathbf{z}-\mu)^*\Gamma^{-1}(\mathbf{z}-\mu)\right] = E\left[tr\left((\mathbf{z}-\mu)^*\Gamma^{-1}(\mathbf{z}-\mu)\right)\right]$$
$$= E\left[tr\left(\Gamma^{-1}(\mathbf{z}-\mu)(\mathbf{z}-\mu)^*\right)\right]$$
$$= tr\left(E\left[\Gamma^{-1}(\mathbf{z}-\mu)(\mathbf{z}-\mu)^*\right]\right)$$
$$= tr\left(\Gamma^{-1}E\left[(\mathbf{z}-\mu)(\mathbf{z}-\mu)^*\right]\right)$$
$$= tr\left(\Gamma^{-1}\Gamma\right)$$
$$= tr(\mathbf{I}_n)$$
$$= n$$

The analogies above between complex and double-real matrices might suggest the result to be $2n$.

However, for real-valued random variables $E\left[(\mathbf{x}-\mu)^*\Sigma^{-1}(\mathbf{x}-\mu)\right] = n$, and the complex case is a superset of the real. So by extension, the complex case must be the same.

But an insight is available into why the value is n, rather than $2n$. Let $\mathbf{x}$ and $\mathbf{y}$ be $n\times1$ real-valued random vectors. Assume their means to be zero, and their variances to be identity matrices (so zero covariance):

$$\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix} \sim \left(\mu = \begin{bmatrix}0_{n\times1}\\0_{n\times1}\end{bmatrix}, \quad \Sigma = \begin{bmatrix}\mathbf{I}_n & 0\\0 & \mathbf{I}_n\end{bmatrix}\right)$$

The quadratic form is:

$$\begin{bmatrix}\mathbf{x}' & \mathbf{y}'\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{x}' & \mathbf{y}'\end{bmatrix}\begin{bmatrix}\mathbf{I}_n & 0\\0 & \mathbf{I}_n\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{x}' & \mathbf{y}'\end{bmatrix}\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix} = \mathbf{x}'\mathbf{x}+\mathbf{y}'\mathbf{y} = \sum_{j=1}^{n}\left(\frac{x_j^2}{1}+\frac{y_j^2}{1}\right)$$

Since the elements have unit variances, the expectation is:

$$E\left[\begin{bmatrix}\mathbf{x}' & \mathbf{y}'\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix}\right] = E\left[\sum_{j=1}^{n}\left(\frac{x_j^2}{1}+\frac{y_j^2}{1}\right)\right] = \sum_{j=1}^{n}\left(\frac{E[x_j^2]}{1}+\frac{E[y_j^2]}{1}\right) = \sum_{j=1}^{n}\left(\frac{1}{1}+\frac{1}{1}\right) = 2n$$

Now let $\mathbf{z}$ be the $n\times1$ complex random vector $\mathbf{x}+i\mathbf{y}$. Since $E[\mathbf{x}+i\mathbf{y}] = 0_{n\times1}$, the variance of $\mathbf{z}$ is:

$$
\begin{aligned}
\Gamma &= Var[\mathbf{z}] \\
&= Var[\mathbf{x} + i\mathbf{y}] \\
&= Var\left[\begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\right] \\
&= E\left[\begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\left\{\begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\right\}^*\right] \\
&= E\left[\begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^*\begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}^*\right] \\
&= \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}E\left[\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^*\right]\begin{bmatrix} \mathbf{I}_n \\ -i\mathbf{I}_n \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}Var\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}\begin{bmatrix} \mathbf{I}_n \\ -i\mathbf{I}_n \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{I}_n & 0 \\ 0 & \mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{I}_n \\ -i\mathbf{I}_n \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \end{bmatrix}\begin{bmatrix} \mathbf{I}_n \\ -i\mathbf{I}_n \end{bmatrix} \\
&= \mathbf{I}_n - i^2\mathbf{I}_n \\
&= 2\mathbf{I}_n
\end{aligned}
$$

The complex quadratic form is:

$$
\mathbf{z}^*\Gamma^{-1}\mathbf{z} = \mathbf{z}^*(2\mathbf{I}_n)^{-1}\mathbf{z} = \frac{\mathbf{z}^*\mathbf{z}}{2} = \sum_{j=1}^{n}\frac{\overline{z}_j z_j}{2} = \sum_{j=1}^{n}\frac{(x_j - iy_j)(x_j + iy_j)}{2} = \sum_{j=1}^{n}\frac{x_j^2 + y_j^2}{1+1} = \frac{1}{2}\begin{bmatrix} \mathbf{x}' & \mathbf{y}' \end{bmatrix}\Sigma^{-1}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}
$$

The complex form is half the real-valued form; hence, its expectation equals *n*. The condensation of

the 2*n* real dimensions into *n* complex ones inverts the order of operations:

$$
\sum_{j=1}^{n}\left(\frac{x_j^2}{1} + \frac{y_j^2}{1}\right) \Rightarrow \sum_{j=1}^{n}\frac{x_j^2 + y_j^2}{1+1}
$$

Within the sigma operator, the sum of two quotients becomes the quotient of two sums. A proof for general variance Γ involves diagonalizing Γ, i.e., that Γ can be eigen-decomposed as $\Gamma = \mathbf{W}\Lambda\mathbf{W}^*$, where Λ is diagonal and $\mathbf{W}\mathbf{W}^* = \mathbf{W}^*\mathbf{W} = \mathbf{I}_n$.[6]

At this point we can derive the standard complex normal distribution. The normal distribution is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$ The standard complex normal random variable is formed from two independent real normal variables whose means equal zero and whose variances equal one half:

$$f_z(z) = \frac{1}{\sqrt{2\pi(1/2)}} e^{-\frac{(x-0)^2}{2(1/2)}} \frac{1}{\sqrt{2\pi(1/2)}} e^{-\frac{(y-0)^2}{2(1/2)}} = \frac{1}{\sqrt{\pi}} e^{-x^2} \frac{1}{\sqrt{\pi}} e^{-y^2} = \frac{1}{\pi} e^{-(x^2+y^2)} = \frac{1}{\pi} e^{-\bar{z}z}$$

The distribution of the $n{\times}1$ standard complex normal random vector is $f_{\mathbf{z}}(\mathbf{z}) = \frac{1}{\pi^n} e^{\mathbf{z}^*\mathbf{z}}$. A vector so distributed has mean $E[\mathbf{z}] = 0_{n \times 1}$ and variance $Var[\mathbf{z}] = E[\mathbf{z}\bar{\mathbf{z}}'] = \mathbf{I}_n$.

6. COMPLEX VARIANCE, PSEUDOVARIANCE, AND AUGMENTED VARIANCE

Section 4 justified the definition of the variance of a complex random vector as:

$$\Gamma = Var[\mathbf{z}] = E\left[(\mathbf{z} - \mu)\overline{(\mathbf{z} - \mu)}'\right] = E\left[(\mathbf{z} - \mu)(\mathbf{z} - \mu)^*\right]$$

The naïve formula differs from this by one critical symbol (prime versus asterisk):

$$C = E\left[(\mathbf{z} - \mu)(\mathbf{z} - \mu)'\right]$$

---

[6] Cf. Appendix C for eigen-decomposition and diagonalization. We believe the insight about commuting sums and quotients to be valuable as an abstraction. But of course, a vector of n independent complex random variables of unit variance translates into a vector of $2n$ independent real random variables of half-unit variance, and $\sum_{j=1}^{2n} \frac{1}{2} = n$. Because of the half-unit real variance, the formula in the next paragraph for the standard complex normal distribution, lacking any factors of two, is simpler than the formula for the standard real normal distribution.

This naïveté leads many to conclude that $Var[i\mathbf{z}] = i^2 Var[\mathbf{z}] = -Var[\mathbf{z}]$, whereas it is actually:[7]

$$Var[i\mathbf{z}] = E\left[i(\mathbf{z}-\mu)\{i(\mathbf{z}-\mu)\}^*\right] = E\left[i(\mathbf{z}-\mu)\bar{i}(\mathbf{z}-\mu)^*\right] = i\bar{i}E\left[(\mathbf{z}-\mu)(\mathbf{z}-\mu)^*\right] = i(-i)Var[\mathbf{z}] = Var[\mathbf{z}]$$

Nevertheless, there is a role for the naïve formula, which reduces to:

$$C = E\left[(\mathbf{z}-\mu)(\mathbf{z}-\mu)'\right] = E\left[(\mathbf{z}-\mu)\overline{(\overline{\mathbf{z}-\mu})}'\right] = E\left[(\mathbf{z}-\mu)(\overline{\mathbf{z}}-\overline{\mu})^*\right] = Cov[\mathbf{z},\overline{\mathbf{z}}]$$

Veeravalli [2006], whose notation we follow, calls C the "relation matrix." The Wikipedia article "Complex normal distribution" calls it the "pseudocovariance matrix." Because of the naïveté that leads many to a false conclusion, we prefer the 'pseudo' terminology (better, "pseudovariance") to something as bland as "relation matrix." However, a useful and non-pejorative concept is what we will call the "augmented variance."

The augmented variance is the variance of the complex random vector $\mathbf{z}$ augmented with its conjugate $\overline{\mathbf{z}}$, i.e., the $2n \times 1$ vector $\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}$. Its expectation is $E\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} E[\mathbf{z}] \\ E[\overline{\mathbf{z}}] \end{bmatrix} = \begin{bmatrix} \mu \\ \overline{\mu} \end{bmatrix}$. And its variance is (for brevity we ignore the mean):

$$Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = E\left[\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}^*\right] = E\left[\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}[\overline{\mathbf{z}}' \quad \mathbf{z}']\right] = \begin{bmatrix} Cov[\mathbf{z},\mathbf{z}] & Cov[\mathbf{z},\overline{\mathbf{z}}] \\ Cov[\overline{\mathbf{z}},\mathbf{z}] & Cov[\overline{\mathbf{z}},\overline{\mathbf{z}}] \end{bmatrix}$$

In two ways this matrix is redundant. First, $Cov[\overline{\mathbf{z}},\overline{\mathbf{z}}] = E[\overline{\mathbf{z}}\mathbf{z}'] = \overline{E[\mathbf{z}\overline{\mathbf{z}}']} = \overline{Cov[\mathbf{z},\mathbf{z}]}$; equivalently, $Var[\overline{\mathbf{z}}] = \overline{Var[\mathbf{z}]} = \overline{\Gamma}$. And second, $Cov[\overline{\mathbf{z}},\mathbf{z}] = E[\overline{\mathbf{z}}\overline{\mathbf{z}}'] = \overline{E[\mathbf{z}\mathbf{z}']} = \overline{Cov[\mathbf{z},\overline{\mathbf{z}}]} = \overline{C}$. Therefore:

$$Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} Cov[\mathbf{z},\mathbf{z}] & Cov[\mathbf{z},\overline{\mathbf{z}}] \\ Cov[\overline{\mathbf{z}},\mathbf{z}] & Cov[\overline{\mathbf{z}},\overline{\mathbf{z}}] \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}$$

---

[7] In general, for any complex scalar α, $Var[\alpha\mathbf{z}] = \alpha\overline{\alpha}Var[\mathbf{z}]$.

As with any valid variance matrix, the augmented variance must be Hermetian. Hence,

$$\begin{bmatrix} \Gamma & C \\ C & \Gamma \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ C & \Gamma \end{bmatrix}^* = \overline{\begin{bmatrix} \Gamma & C \\ C & \Gamma \end{bmatrix}}' = \begin{bmatrix} \overline{\Gamma} & \overline{C} \\ C & \Gamma \end{bmatrix}' = \begin{bmatrix} \Gamma^* & C' \\ C^* & \Gamma' \end{bmatrix}, \text{ from which follow } \Gamma^* = \Gamma \text{ and } C' = C.$$

Moreover, it must be at least NND, if not PD. It is important to note from this that the pseudovariance is an essential part of the augmented **z**; it is possible for two random variables to have the same variance and to covary differently with their conjugates. How a complex random vector covaries with its conjugate is useful information; it is even a parameter of the general complex normal distribution, which we will treat next.

7. THE COMPLEX NORMAL DISTRIBUTION

All the information for deriving the complex normal distribution of $\mathbf{z}_{n \times 1} = \mathbf{x} + i\mathbf{y}$ is contained in the parameters of the real-valued multivariate normal distribution:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(\mu_{2n \times 1} = \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \ \Sigma_{2n \times 2n} = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \right)$$

According to this variance structure, the real and imaginary parts of **z** may covary, as long as the covariance is symmetric: $\Sigma_{\mathbf{yx}} = \Sigma_{\mathbf{xy}}'$. The grand Σ matrix must be symmetric and PD. The probability density function of this multivariate normal is:[8]

$$f_{\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}}(x, y) = \frac{1}{\sqrt{(2\pi)^{2n} |\Sigma|}} e^{-\frac{1}{2} [x' - \mu_{\mathbf{x}}' \ \ y' - \mu_{\mathbf{y}}'] \Sigma^{-1} \begin{bmatrix} x - \mu_{\mathbf{x}} \\ y - \mu_{\mathbf{y}} \end{bmatrix}} = \frac{1}{\pi^n \sqrt{2^{2n} |\Sigma|}} e^{-\frac{1}{2} [x' - \mu_{\mathbf{x}}' \ \ y' - \mu_{\mathbf{y}}'] \Sigma^{-1} \begin{bmatrix} x - \mu_{\mathbf{x}} \\ y - \mu_{\mathbf{y}} \end{bmatrix}}$$

---

[8] As derived briefly by Judge [1988, pp 49f]. Chapter 4 of Johnson [1992] is thorough. To be precise, $|\Sigma|$ under the radical should be $\|\Sigma\|$, the absolute value of the determinant of Σ. However, the determinant of a PD matrix must be positive (cf. Judge [1988, A.14(1)]).

Since $\mathbf{z} = \mathbf{x} + i\mathbf{y}$, the augmented vector is $\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \mathbf{x} + i\mathbf{y} \\ \mathbf{x} - i\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \Xi_n \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$. We will call

$\Xi_n = \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix}$ the augmentation matrix; this linear function of the real-valued vectors produces

the complex vector and its conjugate. An important equation is:

$$\Xi_n \Xi_n^* = \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \\ -i\mathbf{I}_n & i\mathbf{I}_n \end{bmatrix} = \begin{bmatrix} 2\mathbf{I}_n & 0 \\ 0 & 2\mathbf{I}_n \end{bmatrix} = 2\mathbf{I}_{2n}$$

Therefore, Ξ_n has an inverse, viz., one half of its transjugate.

The augmented mean is $E\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \Xi_n \begin{bmatrix} \mu_\mathbf{x} \\ \mu_\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mu_\mathbf{x} + i\mu_\mathbf{y} \\ \mu_\mathbf{x} - i\mu_\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mu \\ \overline{\mu} \end{bmatrix}$. The augmented variance is:

$$\begin{aligned} Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} &= \Xi_n \Sigma \Xi_n^* \\ &= \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \\ -i\mathbf{I}_n & i\mathbf{I}_n \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{\mathbf{xx}} + i\Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{xy}} + i\Sigma_{\mathbf{yy}} \\ \Sigma_{\mathbf{xx}} - i\Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{xy}} - i\Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \\ -i\mathbf{I}_n & i\mathbf{I}_n \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{\mathbf{xx}} + \Sigma_{\mathbf{yy}} - i(\Sigma_{\mathbf{xy}} - \Sigma_{\mathbf{yx}}) & \Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{yy}} + i(\Sigma_{\mathbf{xy}} + \Sigma_{\mathbf{yx}}) \\ \Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{yy}} - i(\Sigma_{\mathbf{xy}} + \Sigma_{\mathbf{yx}}) & \Sigma_{\mathbf{xx}} + \Sigma_{\mathbf{yy}} + i(\Sigma_{\mathbf{xy}} - \Sigma_{\mathbf{yx}}) \end{bmatrix} \\ &= \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix} \end{aligned}$$

And so:

$$Var^{-1}\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}^{-1} = (\Xi_n \Sigma \Xi_n^*)^{-1} = (\Xi_n^*)^{-1} \Sigma^{-1} (\Xi_n)^{-1}$$

This can be reformulated as $\Xi_n^* Var^{-1}\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} \Xi_n = \Xi_n^* \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}^{-1} \Xi_n = \Sigma^{-1}$.

We now work these augmented forms into the probability density function:

$$f_{\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix}}(\mathbf{x},\mathbf{y}) = \frac{1}{\pi^n \sqrt{2^{2n}|\Sigma|}} e^{-\frac{1}{2}[\mathbf{x}'-\mu_\mathbf{x}' \quad \mathbf{y}'-\mu_\mathbf{y}']\Sigma^{-1}\begin{bmatrix}\mathbf{x}-\mu_\mathbf{x}\\\mathbf{y}-\mu_\mathbf{y}\end{bmatrix}}$$

$$= \frac{1}{\pi^n \sqrt{|2I_{2n}||\Sigma|}} e^{-\frac{1}{2}[\mathbf{x}'-\mu_\mathbf{x}' \quad \mathbf{y}'-\mu_\mathbf{y}']\Xi_n^* Var^{-1}\begin{bmatrix}\mathbf{z}\\\bar{\mathbf{z}}\end{bmatrix}\Xi_n\begin{bmatrix}\mathbf{x}-\mu_\mathbf{x}\\\mathbf{y}-\mu_\mathbf{y}\end{bmatrix}}$$

$$= \frac{1}{\pi^n \sqrt{|\Xi_n\Xi_n^*||\Sigma|}} e^{-\frac{1}{2}\left(\Xi_n\begin{bmatrix}\mathbf{x}-\mu_\mathbf{x}\\\mathbf{y}-\mu_\mathbf{y}\end{bmatrix}\right)^* Var^{-1}\begin{bmatrix}\mathbf{z}\\\bar{\mathbf{z}}\end{bmatrix}\left(\Xi_n\begin{bmatrix}\mathbf{x}-\mu_\mathbf{x}\\\mathbf{y}-\mu_\mathbf{y}\end{bmatrix}\right)}$$

$$= \frac{1}{\pi^n \sqrt{|\Xi_n\Sigma\Xi_n^*|}} e^{-\frac{1}{2}[\bar{\mathbf{z}}'-\bar{\mu}' \quad \mathbf{z}'-\mu']Var^{-1}\begin{bmatrix}\mathbf{z}\\\bar{\mathbf{z}}\end{bmatrix}\begin{bmatrix}\mathbf{z}-\mu\\\bar{\mathbf{z}}-\bar{\mu}\end{bmatrix}}$$

$$= \frac{1}{\pi^n \sqrt{\left|Var\begin{bmatrix}\mathbf{z}\\\bar{\mathbf{z}}\end{bmatrix}\right|}} e^{-\frac{1}{2}[\bar{\mathbf{z}}'-\bar{\mu}' \quad \mathbf{z}'-\mu']Var^{-1}\begin{bmatrix}\mathbf{z}\\\bar{\mathbf{z}}\end{bmatrix}\begin{bmatrix}\mathbf{z}-\mu\\\bar{\mathbf{z}}-\bar{\mu}\end{bmatrix}}$$

However, this is not quite the density function of **z**, since the differential volume has not been considered. The correct formula is $f_\mathbf{z}(\mathbf{z})dV_\mathbf{z} = f_{\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix}}(\mathbf{x},\mathbf{y})dV_{xy}$. The differential volume in the xy

coordinates is $dV_{xy} = \left|\prod_{j=1}^n dx_j dy_j\right|$. A change of dx_j entails an equal change in the real part of dz_j,

even as a change of dy_j entails an equal change in the imaginary part of dz_j. Accordingly,

$dV_\mathbf{z} = \left|\prod_{j=1}^n dx_j(i \cdot dy_j)\right| = \left|i^n \prod_{j=1}^n dx_j dy_j\right| = |i^n|\left|\prod_{j=1}^n dx_j dy_j\right| = 1 \cdot \left|\prod_{j=1}^n dx_j dy_j\right| = dV_{xy}$. It so happens that

Ξ_n does not distort volume; but this had to be demonstrated.[9]

So finally, the probability density function of the complex random vector **z** is:

---

[9] This will be abstruse to some actuaries. However, the integration theory is implicit in the change-of-variables technique outlined in Hogg [1984, pp 42-46]. That the $n \times n$ determinant represents "the volume function of an *n*-dimensional parallelepiped" is beautifully explained in Chapter 4 of Schneider [1973].

$$f_{\mathbf{z}}(z) = f_{\mathbf{z}}(z) \cdot 1 = f_{\mathbf{z}}(z)\frac{dV_{\mathbf{z}}}{dV_{xy}} = f_{\begin{bmatrix}\mathbf{x}\\\mathbf{y}\end{bmatrix}}(x, y)$$

$$= \frac{1}{\pi^n \sqrt{\left|Var\begin{bmatrix}\mathbf{z}\\\overline{\mathbf{z}}\end{bmatrix}\right|}} e^{-\frac{1}{2}\begin{bmatrix}\overline{z'}-\overline{\mu'} & z'-\mu'\end{bmatrix}Var^{-1}\begin{bmatrix}\mathbf{z}\\\overline{\mathbf{z}}\end{bmatrix}\begin{bmatrix}z-\mu\\\overline{z}-\overline{\mu}\end{bmatrix}}$$

$$= \frac{1}{\pi^n \sqrt{\left|\begin{matrix}\Gamma & C\\\overline{C} & \overline{\Gamma}\end{matrix}\right|}} e^{-\frac{1}{2}\begin{bmatrix}\overline{z'}-\overline{\mu'} & z'-\mu'\end{bmatrix}\begin{bmatrix}\Gamma & C\\\overline{C} & \overline{\Gamma}\end{bmatrix}^{-1}\begin{bmatrix}z-\mu\\\overline{z}-\overline{\mu}\end{bmatrix}}$$

This formula is equivalent to the one found in the Wikipedia article "Complex normal distribution."

Although $\left|\Gamma\right|\left|\overline{\Gamma} - \overline{C}\Gamma^{-1}C\right|$ appears within the radical of that article's formula, it can be shown that

$\left|\begin{matrix}\Gamma & C\\\overline{C} & \overline{\Gamma}\end{matrix}\right| = \left|\Gamma\right|\left|\overline{\Gamma} - \overline{C}\Gamma^{-1}C\right|$. As far as allowable parameters are concerned, μ may be any complex

vector. $\left|\begin{matrix}\Gamma & C\\\overline{C} & \overline{\Gamma}\end{matrix}\right|$ is allowed if and only if $\Xi_n^*\begin{bmatrix}\Gamma & C\\\overline{C} & \overline{\Gamma}\end{bmatrix}\Xi_n = 4\Sigma$ is real-valued and PD.

Veeravalli [2006] defines a "proper" complex variable as one whose pseudo[co]variance matrix is $0_{n\times n}$. Inserting zero for C into the formula, we derive the probability density function of a proper complex random variable whose variance is Γ:

$$f_{\mathbf{z}}(z) = \frac{1}{\pi^n \sqrt{\left| \begin{matrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{matrix} \right|}} e^{-\frac{1}{2}\left[\overline{z}' - \overline{\mu}' \quad z' - \mu' \right] \left[\begin{matrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{matrix} \right]^{-1} \left[\begin{matrix} z - \mu \\ \overline{z} - \overline{\mu} \end{matrix} \right]}$$

$$= \frac{1}{\pi^n \sqrt{\left| \begin{matrix} \Gamma & 0 \\ 0 & \overline{\Gamma} \end{matrix} \right|}} e^{-\frac{1}{2}\left[\overline{z}' - \overline{\mu}' \quad z' - \mu' \right] \left[\begin{matrix} \Gamma & 0 \\ 0 & \overline{\Gamma} \end{matrix} \right]^{-1} \left[\begin{matrix} z - \mu \\ \overline{z} - \overline{\mu} \end{matrix} \right]}$$

$$= \frac{1}{\pi^n \sqrt{|\Gamma||\overline{\Gamma}|}} e^{-\frac{1}{2}\left\{ \left(\overline{z}' - \overline{\mu}' \right) \Gamma^{-1} (z - \mu) + (z' - \mu') \overline{\Gamma}^{-1} \left(\overline{z} - \overline{\mu} \right) \right\}}$$

$$= \frac{1}{\pi^n \sqrt{|\Gamma||\overline{\Gamma}|}} e^{-\frac{1}{2}\left\{ \left(\overline{z}' - \overline{\mu}' \right) \Gamma^{-1} (z - \mu) + \overline{\left(\overline{z}' - \overline{\mu}' \right) \Gamma^{-1} (z' - \mu')} \right\}}$$

$$= \frac{1}{\pi^n \sqrt{|\Gamma|^2}} e^{-\frac{1}{2}\left\{ \left(\overline{z}' - \overline{\mu}' \right) \Gamma^{-1} (z - \mu) + \left(\overline{z}' - \overline{\mu}' \right) \Gamma^{-1} (z - \mu) \right\}}$$

$$= \frac{1}{\pi^n |\Gamma|} e^{-\left(\overline{z}' - \overline{\mu}' \right) \Gamma^{-1} (z - \mu)}$$

$$= \frac{1}{\pi^n |\Gamma|} e^{-(z - \mu)^* \Gamma^{-1} (z - \mu)}$$

The transformations in the last several lines rely on the fact that Γ is Hermetian and PD. Now the standard complex random vector is a proper complex random vector with mean zero and variance I_n. Therefore, in confirmation of Section 5, its density function is $\frac{1}{\pi^n} e^{-z^* z}$.

8. THE COMPLEX LOGNORMAL RANDOM VECTOR AND ITS MOMENTS

A complex lognormal random vector is the elementwise exponentiation of a complex normal random vector: $\mathbf{w}_{n \times 1} = e^{\mathbf{z}_{n \times 1}}$. Its conjugate also is lognormal, since $\overline{\mathbf{w}} = \overline{e^{\mathbf{z}}} = e^{\overline{z}}$. Deriving the probability density function of $\mathbf{w}$ is precluded by the fact that $e : z \to w$ is many-to-one. Specifically, $e^z = w = e^{z + i(2\pi k)}$ for any integer k. So unlike the real-valued lognormal random variable, whose density function can be found in Klugman [1998, §A.4.11], an analytic form for the

complex lognormal density is not available. However, even for the real-valued lognormal the density function is of little value; its moments are commonly derived from the moment generating function of the normal variable on which it is based. So too, the moment generating function of the complex normal random vector is available for deriving the lognormal moments.

We hereby define the moment generating function of the complex $n \times 1$ random vector $\mathbf{z}$ as $M_{\mathbf{z}}(s_{n \times 1}, t_{n \times 1}) = E[e^{s'\mathbf{z} + t'\overline{\mathbf{z}}}]$. Since this definition may differ from other definitions in the sparse literature, we should justify it. First, because we will take derivatives of this function with respect to s and t, the function must be differentiable. This demands simple transposition in the linear combination, i.e., $s'\mathbf{z} + t'\overline{\mathbf{z}}$ rather than the transjugation $s^*\mathbf{z} + t^*\overline{\mathbf{z}}$. For transjugation would involve derivatives of the form $\dfrac{d\overline{s}}{ds}$, which do not exist, as they violate the Cauchy-Riemann condition.[10] Second, even though moments of $\overline{\mathbf{z}}$ are conjugates of moments of $\mathbf{z}$, we will need second-order moments involving both $\mathbf{z}$ and $\overline{\mathbf{z}}$. For this reason both terms must be in the exponent of the moment generating function.

---

[10] Cf. Appendix D.1.3 of Havil [2003]. Express $f(z = x + iy)$ in terms of real-valued functions, i.e., as $u(x, y) + i \cdot v(x, y)$. The derivative is based on the matrix of real-valued partial derivatives $\begin{bmatrix} \partial u/\partial x & \partial v/\partial x \\ \partial u/\partial y & \partial v/\partial y \end{bmatrix}$. For the derivative to be the same in both directions, the Cauchy-Riemann condition must hold, viz., that $\partial u/\partial x = \partial v/\partial y$ and $\partial v/\partial x = -\partial u/\partial y$. But for $f(z) = \overline{z} = x - iy$, the partial-derivative matrix is $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$; hence $\partial u/\partial x \neq \partial v/\partial y$. The Cauchy-Riemann condition becomes intuitive when one regards a valid complex derivative as a double-real 2×2 matrix (Section 3). Compare this with $f(z) = z = x + iy$, whose matrix is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, which represents the complex number 1.

We start with terminology from Section 7, viz., that $\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \mathbf{x} + i\mathbf{y} \\ \mathbf{x} - i\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \Xi_n \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ and

that $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N\left(\mu_{2n\times 1} = \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \Sigma_{2n\times 2n} = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \right)$. According to Appendix A, the moment

generating function of the real-valued normal random vector $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}$ is:

$$M_{\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}}\left(\begin{bmatrix} a \\ b \end{bmatrix} \right) = E\left[e^{[a' \ b'] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}} \right] = e^{[a' \ b'] \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix} + \frac{1}{2}[a' \ b'] \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}}$$

Consequently:

$$
\begin{aligned}
M_{\mathbf{z}}(s, t) &= E\left[e^{s'\mathbf{z} + t'\overline{\mathbf{z}}} \right] \\
&= E\left[e^{[s' \ t'] \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}} \right] \\
&= E\left[e^{[s'+t' \ i(s'-t')] \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}} \right] \\
&= M_{\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}}\left(\begin{bmatrix} s+t \\ i(s-t) \end{bmatrix} \right) \\
&= e^{[(s+t)' \ i(s-t)'] \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix} + \frac{1}{2}[(s+t)' \ i(s-t)'] \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} s+t \\ i(s-t) \end{bmatrix}}
\end{aligned}
$$

It is so that we could invoke it here that Appendix B went to the trouble of proving that complex values are allowed in this moment generating function.

But in two ways we can simplify this expression. First:

$$\begin{bmatrix} (s+t)' & i(s-t)' \end{bmatrix} \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix} = (s+t)'\mu_{\mathbf{x}} + i(s-t)'\mu_{\mathbf{y}} = s'(\mu_{\mathbf{x}} + i\mu_{\mathbf{y}}) + t'(\mu_{\mathbf{x}} - i\mu_{\mathbf{y}}) = s'\mu_{\mathbf{z}} + t'\overline{\mu}_{\mathbf{z}}$$

And second, again from Section 7, $Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix} = \Xi_n \Sigma_{2n \times 2n} \Xi_n^* = \Xi_n \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \Xi_n^*$, or

equivalently, $\dfrac{\Xi_n^*}{2} \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix} \dfrac{\Xi_n}{2} = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}$. Hence:

$$\begin{bmatrix} (s+t)' & i(s-t)' \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} s+t \\ i(s-t) \end{bmatrix} = \begin{bmatrix} (s+t)' & i(s-t)' \end{bmatrix} \dfrac{\Xi_n^*}{2} \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix} \dfrac{\Xi_n}{2} \begin{bmatrix} s+t \\ i(s-t) \end{bmatrix}$$

On the right side, $\dfrac{\Xi_n}{2} \begin{bmatrix} s+t \\ i(s-t) \end{bmatrix} = \dfrac{1}{2} \begin{bmatrix} I_n & iI_n \\ I_n & -iI_n \end{bmatrix} \begin{bmatrix} s+t \\ i(s-t) \end{bmatrix} = \dfrac{1}{2} \begin{bmatrix} (s+t)-(s-t) \\ (s+t)+(s-t) \end{bmatrix} = \begin{bmatrix} t \\ s \end{bmatrix}$. And on the

left side, $\begin{bmatrix} (s+t)' & i(s-t)' \end{bmatrix} \dfrac{\Xi_n^*}{2} = \dfrac{1}{2} \begin{bmatrix} (s+t)' & i(s-t)' \end{bmatrix} \begin{bmatrix} I_n & I_n \\ -iI_n & iI_n \end{bmatrix} = \begin{bmatrix} s' & t' \end{bmatrix}$. So:

$$\begin{bmatrix} (s+t)' & i(s-t)' \end{bmatrix} \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} s+t \\ i(s-t) \end{bmatrix} = \begin{bmatrix} s' & t' \end{bmatrix} \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix} \begin{bmatrix} t \\ s \end{bmatrix}$$

And the simplified expression, based on the mean and the augmented variance of $\mathbf{z}$, is:

$$M_{\mathbf{z}}(s,t) = E\left[e^{s'\mathbf{z} + t'\overline{\mathbf{z}}} \right]$$

$$= e^{s'\mu_{\mathbf{z}} + t'\overline{\mu}_{\mathbf{z}} + \frac{1}{2}\begin{bmatrix} s' & t' \end{bmatrix} \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix} \begin{bmatrix} t \\ s \end{bmatrix}}$$

$$= e^{s'\mu + t'\overline{\mu} + \frac{1}{2}(s'\Gamma t + s'Cs + t'\overline{C}t + t'\overline{\Gamma}s)}$$

As in Appendix A, let e_j denote the jth unit vector of $\mathfrak{R}^n$, or even better, of C^n. Then:

$$E\left[e^{\mathbf{z}_j} \right] = M_{\mathbf{z}}(e_j, 0) = e^{\mu_j + \frac{1}{2}C_{jj}}$$

$$E\left[\overline{e^{\mathbf{z}_j}} \right] = E\left[e^{\overline{\mathbf{z}}_j} \right] = M_{\mathbf{z}}(0, e_j) = e^{\overline{\mu}_j + \frac{1}{2}\overline{C}_{jj}} = \overline{E\left[e^{\mathbf{z}_j} \right]}$$

Moreover:

$$E\left[e^{\mathbf{z}_j} e^{\mathbf{z}_k} \right] = E\left[e^{\mathbf{z}_j + \mathbf{z}_k} \right] = M_{\mathbf{z}}(e_j + e_k, 0) = e^{\mu_j + \mu_k + \frac{1}{2}(C_{jj} + C_{jk} + C_{kj} + C_{kk})}$$

According to Section 6, C is symmetric ($\mathbf{C}' = \mathbf{C}$). This and further simplification leads to:

$$E\left[e^{\mathbf{z}_j}e^{\mathbf{z}_k}\right]=e^{\mu_j+\mu_k+\frac{1}{2}\left(C_{jj}+C_{jk}+C_{kj}+C_{kk}\right)}=e^{\left(\mu_j+\frac{1}{2}C_{jj}\right)+\left(\mu_k+\frac{1}{2}C_{kk}\right)+\frac{1}{2}\left(C_{jk}+C_{jk}\right)}=E\left[e^{\mathbf{z}_j}\right]E\left[e^{\mathbf{z}_k}\right]\cdot e^{C_{jk}}$$

Hence, mindful of the transjugation (*) in the definition of complex covariance, we have:

$$Cov\left[e^{\mathbf{z}_j},\overline{e^{\mathbf{z}_k}}\right]=E\left[e^{\mathbf{z}_j}\overline{e^{\mathbf{z}_k}}\right]-E\left[e^{\mathbf{z}_j}\right]E\left[\overline{e^{\mathbf{z}_k}}\right]=E\left[e^{\mathbf{z}_j}e^{\mathbf{z}_k}\right]-E\left[e^{\mathbf{z}_j}\right]E\left[e^{\mathbf{z}_k}\right]=E\left[e^{\mathbf{z}_j}\right]E\left[e^{\mathbf{z}_k}\right]\cdot\left(e^{C_{jk}}-1\right)$$

In terms of $\mathbf{w}=e^{\mathbf{z}}$ this translates as $Cov[\mathbf{w},\overline{\mathbf{w}}]=\left(E[\mathbf{w}]E[\mathbf{w}]'\right)\circ\left(e^C-1_{n\times n}\right)$, in which the '∘' operator represents elementwise multiplication.[11] So too:

$$Cov\left[\overline{e^{\mathbf{z}_j}},e^{\mathbf{z}_k}\right]=\overline{Cov\left[e^{\mathbf{z}_j},\overline{e^{\mathbf{z}_k}}\right]}=E\left[\overline{e^{\mathbf{z}_j}}\right]E\left[e^{\mathbf{z}_k}\right]\cdot\left(e^{\overline{C}_{jk}}-1\right)$$

This translates as $Cov[\overline{\mathbf{w}},\mathbf{w}]=\left(E[\overline{\mathbf{w}}]E[\overline{\mathbf{w}}]'\right)\circ\left(e^{\overline{C}}-1_{n\times n}\right)$.

The remaining combination is the mixed form $E\left[e^{\mathbf{z}_j}\overline{e^{\mathbf{z}_k}}\right]$:

$$E\left[e^{\mathbf{z}_j}\overline{e^{\mathbf{z}_k}}\right]=E\left[e^{\mathbf{z}_j}e^{\overline{\mathbf{z}}_k}\right]=E\left[e^{\mathbf{z}_j+\overline{\mathbf{z}}_k}\right]=M_{\mathbf{z}}\left(e_j,e_k\right)=e^{\mu_j+\overline{\mu}_k+\frac{1}{2}\left(\Gamma_{jk}+C_{jj}+\overline{C}_{kk}+\overline{\Gamma}_{kj}\right)}$$

Since Γ is Hermetian, $\overline{\Gamma}_{kj}=\overline{\Gamma}'_{jk}=\Gamma^*_{jk}=\Gamma_{jk}$. Hence:

$$E\left[e^{\mathbf{z}_j}\overline{e^{\mathbf{z}_k}}\right]=e^{\mu_j+\overline{\mu}_k+\frac{1}{2}\left(\Gamma_{jk}+C_{jj}+\overline{C}_{kk}+\overline{\Gamma}_{kj}\right)}=e^{\left(\mu_j+\frac{1}{2}C_{jj}\right)+\left(\overline{\mu}_k+\frac{1}{2}\overline{C}_{kk}\right)+\frac{1}{2}\left(\Gamma_{jk}+\Gamma_{jk}\right)}=E\left[e^{\mathbf{z}_j}\right]E\left[\overline{e^{\mathbf{z}_k}}\right]\cdot e^{\Gamma_{jk}}$$

Therefore, $Cov\left[e^{\mathbf{z}_j},e^{\mathbf{z}_k}\right]=E\left[e^{\mathbf{z}_j}\overline{e^{\mathbf{z}_k}}\right]-E\left[e^{\mathbf{z}_j}\right]E\left[\overline{e^{\mathbf{z}_k}}\right]=E\left[e^{\mathbf{z}_j}\right]E\left[\overline{e^{\mathbf{z}_k}}\right]\cdot\left(e^{\Gamma_{jk}}-1\right)$, which translates as

$Cov[\mathbf{w},\mathbf{w}]=\left(E[\mathbf{w}]E[\overline{\mathbf{w}}]'\right)\circ\left(e^\Gamma-1_{n\times n}\right)$. By conjugation, $Cov\left[\overline{e^{\mathbf{z}_j}},\overline{e^{\mathbf{z}_k}}\right]=E\left[\overline{e^{\mathbf{z}_j}}\right]E\left[e^{\mathbf{z}_k}\right]\cdot\left(e^{\overline{\Gamma}_{jk}}-1\right)$,

which translates as $Cov[\overline{\mathbf{w}},\overline{\mathbf{w}}]=\left(E[\overline{\mathbf{w}}]E[\mathbf{w}]'\right)\circ\left(e^{\overline{\Gamma}}-1_{n\times n}\right)$.

---

[11] Elementwise multiplication is formally known as the Hadamard, or Hadamard-Schur, product, of which we will make use in Appendices A and C. Cf. Million [2007].

We conclude this section by expressing it all in terms of $\mathbf{w}_{n\times 1} = e^{\mathbf{z}_{n\times 1}}$. Let z be complex normal with

mean μ and augmented variance $Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}$. And let D be the $n\times 1$ vector consisting of the

main diagonal of C. Then $\overline{D}$ is the vectorization of the diagonal of $\overline{C}$. So the augmented mean of

$\mathbf{w}$ is $E\begin{bmatrix} \mathbf{w} \\ \overline{\mathbf{w}} \end{bmatrix} = \begin{bmatrix} e^{\mu + D/2} \\ e^{\overline{\mu} + \overline{D}/2} \end{bmatrix}$. And the augmented variance of $\mathbf{w}$ is:

$$Var\begin{bmatrix} \mathbf{w} \\ \overline{\mathbf{w}} \end{bmatrix} = \begin{bmatrix} Cov[\mathbf{w}, \mathbf{w}] & Cov[\mathbf{w}, \overline{\mathbf{w}}] \\ Cov[\overline{\mathbf{w}}, \mathbf{w}] & Cov[\overline{\mathbf{w}}, \overline{\mathbf{w}}] \end{bmatrix}$$

$$= \begin{bmatrix} \left(E[\mathbf{w}]E[\overline{\mathbf{w}}]'\right) \circ \left(e^{\Gamma} - 1_{n\times n}\right) & \left(E[\mathbf{w}]E[\mathbf{w}]'\right) \circ \left(e^{C} - 1_{n\times n}\right) \\ \left(E[\overline{\mathbf{w}}]E[\overline{\mathbf{w}}]'\right) \circ \left(e^{\overline{C}} - 1_{n\times n}\right) & \left(E[\overline{\mathbf{w}}]E[\mathbf{w}]'\right) \circ \left(e^{\overline{\Gamma}} - 1_{n\times n}\right) \end{bmatrix}$$

$$= \left(\begin{matrix} E[\mathbf{w}]E[\overline{\mathbf{w}}]' & E[\mathbf{w}]E[\mathbf{w}]' \\ E[\overline{\mathbf{w}}]E[\overline{\mathbf{w}}]' & E[\overline{\mathbf{w}}]E[\mathbf{w}]' \end{matrix}\right) \circ \left(e^{\begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}} - 1_{2n\times 2n}\right)$$

$$= \left(E\begin{bmatrix} \mathbf{w} \\ \overline{\mathbf{w}} \end{bmatrix} E[\overline{\mathbf{w}} \quad \mathbf{w}]'\right) \circ \left(e^{Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}} - 1_{2n\times 2n}\right)$$

$$= \left(E\begin{bmatrix} \mathbf{w} \\ \overline{\mathbf{w}} \end{bmatrix} E\begin{bmatrix} \mathbf{w} \\ \overline{\mathbf{w}} \end{bmatrix}^*\right) \circ \left(e^{Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}} - 1_{2n\times 2n}\right)$$

Scaling all the lognormal means to unity (or setting $\mu = -D/2$), we can say that the coefficient-of-

lognormal-augmented-variation matrix equals $e^{Var\begin{bmatrix} \mathbf{z} \\ \overline{\mathbf{z}} \end{bmatrix}} - 1_{2n\times 2n}$, which is analogous with the well-

known coefficient of lognormal variation $e^{\sigma^2} - 1$.

9. THE COMPLEX LOGNORMAL RANDOM VARIABLE

The previous section derived the augmented mean and variance of the lognormal random vector;

this section provides some intuition into it. The complex lognormal random variable, or scalar,

derives from the real-valued normal bivariate $\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{bmatrix} \right)$. Zero is not much of a

restriction; since $e^{CN(\mu,\,V)} = e^{\mu + CN(0,\,V)} = e^\mu \circ e^{CN(0,\,V)}$, the normal mean affects only the scale of the

lognormal. The variance is written in correlation form, where $-1 \le \rho \le 1$. As usual, $0 < \sigma, \tau < \infty$.

Define $\begin{bmatrix} Z \\ \overline{Z} \end{bmatrix} = \Xi_1 \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X + iY \\ X - iY \end{bmatrix}$. Its mean is zero, and according to Section 7 its

variance (the augmented variance) is:

$$
\begin{aligned}
Var\begin{bmatrix} Z \\ \overline{Z} \end{bmatrix} &= \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \begin{bmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2 + \tau^2 - i(\rho\sigma\tau - \rho\sigma\tau) & \sigma^2 - \tau^2 + i(\rho\sigma\tau + \rho\sigma\tau) \\ \sigma^2 - \tau^2 - i(\rho\sigma\tau + \rho\sigma\tau) & \sigma^2 + \tau^2 + i(\rho\sigma\tau - \rho\sigma\tau) \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2 + \tau^2 & \sigma^2 - \tau^2 + 2i\rho\sigma\tau \\ \sigma^2 - \tau^2 - 2i\rho\sigma\tau & \sigma^2 + \tau^2 \end{bmatrix}
\end{aligned}
$$

We will say little about non-zero correlation ($\rho \ne 0$); but at this point a digression on complex

correlation is apt. The coefficient of correlation between Z and its conjugate is:

$$
\rho_{Z\overline{Z}} = \frac{\sigma^2 - \tau^2 + 2i\rho\sigma\tau}{\sigma^2 + \tau^2} = \overline{\frac{\sigma^2 - \tau^2 - 2i\rho\sigma\tau}{\sigma^2 + \tau^2}} = \overline{\rho_{\overline{Z}Z}}
$$

As a form of covariance, correlation is Hermetian. Moreover:

$$
0 \le \rho_{Z\overline{Z}} \overline{\rho_{Z\overline{Z}}} = \rho_{Z\overline{Z}} \rho_{\overline{Z}Z} = \frac{\left(\sigma^2 - \tau^2\right)^2 + 4\rho^2\sigma^2\tau^2}{\left(\sigma^2 + \tau^2\right)^2} \le \frac{\left(\sigma^2 - \tau^2\right)^2 + 4(1)\sigma^2\tau^2}{\left(\sigma^2 + \tau^2\right)^2} = \frac{\left(\sigma^2 + \tau^2\right)^2}{\left(\sigma^2 + \tau^2\right)^2} = 1
$$

So, the magnitude of complex correlation is not greater than unity. The imaginary part of the

correlation is zero unless some correlation exists between the real and imaginary parts of the

underlying bivariate. More interesting are the two limits: $\lim_{\tau^2 \to 0^+} \rho_{Z\overline{Z}} = 1$ and $\lim_{\sigma^2 \to 0^+} \rho_{Z\overline{Z}} = -1$. In the

first case, $\overline{Z} \to Z$ in a statistical sense, and the correlation approaches one. In the second case,

$\overline{Z} \to -Z$, and the correlation approaches negative one.

Now if $W = e^Z$, by the formulas of Section 8, $E[W] = e^{0 + (\sigma^2 - \tau^2 + 2i\rho\sigma\tau)/2} = e^{(\sigma^2 - \tau^2)/2} \cdot e^{i\rho\sigma\tau}$ and

$E[\overline{W}] = e^{(\sigma^2 - \tau^2)/2} \cdot e^{-i\rho\sigma\tau}$. And the augmented variance is:

$$Var\begin{bmatrix} W \\ \overline{W} \end{bmatrix} = \left(E\begin{bmatrix} W \\ \overline{W} \end{bmatrix} E\begin{bmatrix} W \\ \overline{W} \end{bmatrix}^* \right) \circ \left(e^{Var\begin{bmatrix} Z \\ \overline{Z} \end{bmatrix}} - 1_{2\times 2} \right)$$

$$= \left(\begin{bmatrix} e^{(\sigma^2 - \tau^2)/2} \cdot e^{i\rho\sigma\tau} \\ e^{(\sigma^2 - \tau^2)/2} \cdot e^{-i\rho\sigma\tau} \end{bmatrix} \begin{bmatrix} e^{(\sigma^2 - \tau^2)/2} \cdot e^{-i\rho\sigma\tau} & e^{(\sigma^2 - \tau^2)/2} \cdot e^{i\rho\sigma\tau} \end{bmatrix} \right) \circ \left(e^{\begin{bmatrix} \sigma^2 + \tau^2 & \sigma^2 - \tau^2 + i\cdot 2\rho\sigma\tau \\ \sigma^2 - \tau^2 - i\cdot 2\rho\sigma\tau & \sigma^2 + \tau^2 \end{bmatrix}} - 1_{2\times 2} \right)$$

$$= e^{\sigma^2 - \tau^2} \begin{bmatrix} 1 & e^{2i\rho\sigma\tau} \\ e^{-2i\rho\sigma\tau} & 1 \end{bmatrix} \circ \begin{bmatrix} e^{\sigma^2 + \tau^2} - 1 & e^{\sigma^2 - \tau^2 + 2i\rho\sigma\tau} - 1 \\ e^{\sigma^2 - \tau^2 - 2i\rho\sigma\tau} - 1 & e^{\sigma^2 + \tau^2} - 1 \end{bmatrix}$$

$$= e^{\sigma^2 - \tau^2} \begin{bmatrix} e^{\sigma^2 + \tau^2} - 1 & e^{\sigma^2 - \tau^2 + 4i\rho\sigma\tau} - e^{2i\rho\sigma\tau} \\ e^{\sigma^2 - \tau^2 - 4i\rho\sigma\tau} - e^{-2i\rho\sigma\tau} & e^{\sigma^2 + \tau^2} - 1 \end{bmatrix}$$

$$= \begin{bmatrix} e^{2\sigma^2} - e^{\sigma^2 - \tau^2} & e^{2\sigma^2 - 2\tau^2 + 4i\rho\sigma\tau} - e^{\sigma^2 - \tau^2} \cdot e^{2i\rho\sigma\tau} \\ e^{2\sigma^2 - 2\tau^2 - 4i\rho\sigma\tau} - e^{\sigma^2 - \tau^2} \cdot e^{-2i\rho\sigma\tau} & e^{2\sigma^2} - e^{\sigma^2 - \tau^2} \end{bmatrix}$$

In the first case above, as $\tau^2 \to 0^+$, $E\begin{bmatrix} W \\ \overline{W} \end{bmatrix} \to e^{\sigma^2/2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $Var\begin{bmatrix} W \\ \overline{W} \end{bmatrix} \to e^{\sigma^2}\left(e^{\sigma^2} - 1\right)\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Since

the complex part Y becomes probability-limited to its mean of zero, the complex lognormal

degenerates to the real-valued $W = e^X$. The limiting result is oblivious to the underlying correlation

ϱ, since $\overline{W} \to W$.

In the second case, as $\sigma^2 \to 0^+$, $E\left[\dfrac{W}{\overline{W}}\right] \to e^{-\tau^2/2}\begin{bmatrix}1\\1\end{bmatrix}$ and $Var\left[\dfrac{W}{\overline{W}}\right] \to e^{-\tau^2}\begin{bmatrix}e^{\tau^2}-1 & e^{-\tau^2}-1\\ e^{-\tau^2}-1 & e^{\tau^2}-1\end{bmatrix}$. As

in the first case, both $E[\overline{W}] = E[W]$ and the underlying correlation ϱ has disappeared. Nevertheless, the variance shows W and its conjugate to differ; in fact, their correlation is the real-valued

$\rho_{W\overline{W}} = \left(e^{-\tau^2}-1\right)\big/\left(e^{\tau^2}-1\right) = -e^{-\tau^2} = -E[W].$ Since $\tau^2 > 0$, $-1 < \rho_{W\overline{W}} < 0$ and

$0 < E[\overline{W}] = E[W] < 1.$

Both these cases are understandable from the "geometry" of $W = e^{Z} = e^{X+iY} = e^{X}e^{iY}$. The complex exponential function is the basis of polar coordinates; e^{X} is the magnitude of W, and Y is the angle of W in radians counterclockwise from the real axis of the complex plane. Imagine a canon whose angle and range can be set. In the first case, the angle is fixed at zero, but the range is variable. This makes for a lognormal distribution along the positive real axis. In the second case, the canon's angle varies, but its range is fixed at $e^{0} = 1$. This makes all the shots to land on the complex unit circle; hence, their mean lies within the circle, i.e., $|E[W]| < 1$. Moreover, the symmetry of Y as $N(0, \tau^2)$-distributed guarantees $E[W]$ to fall on the real axis, or $-1 < E[W] < 1$. Furthermore, since the normal density function strictly decreases in both directions from the mean, more shots land to the right of the imaginary axis than to the left, so $0 < E[W] = e^{-\tau^2} < 1$. A "right-handed" canon, or a canon whose angle is measured clockwise from the real axis, fires $\overline{W} = e^{X}e^{-iY}$ shots.

A shot from an unrestricted canon will "almost surely" not land on the real axis.[12] If we desire

negative values from the complex lognormal random variable, as a practical matter we must extract

them from its real or complex parts, e.g., $U = \text{Re}(W)$. One can see in the second case, that as τ^2

grows larger, so too grows larger the probability that $U < 0$. As $\tau^2 \to \infty$, the probability

approaches one half. In the limit, the shots are uniformly distributed around the complex unit

circle. In this specialized case ($\sigma^2 \to 0^+$ and $\tau^2 \to \infty$), the distribution of $U = \text{Re}(W)$ is

$$f_U(u) = \frac{1}{\pi\sqrt{1-u^2}}, \text{ for } -1 \le u \le 1.[13]$$

This suggests a third case, in which $\tau^2 \to \infty$ while σ^2 remains at some positive amount. An

intriguing feature of complex variables is that infinite variance in Y leads to a uniform distribution of

e^{iY}.[14] So if $W = e^Z = e^X e^{iY}$, $U = \text{Re}(W) = e^X \cos Y$ will be something of a reflected lognormal;

both its tails will be as heavy as the lognormal's.[15] In this case:

$$E\begin{bmatrix} W \\ \overline{W} \end{bmatrix} = \lim_{\tau^2 \to \infty} \begin{bmatrix} e^{(\sigma^2 - \tau^2 + 2i\rho\sigma\tau)/2} \\ e^{(\sigma^2 - \tau^2 + 2i\rho\sigma\tau)/2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Var\begin{bmatrix} W \\ \overline{W} \end{bmatrix} = \lim_{\tau^2 \to \infty} \begin{bmatrix} e^{2\sigma^2} - e^{\sigma^2 - \tau^2} & e^{2\sigma^2 - 2\tau^2 + 4i\rho\sigma\tau} - e^{\sigma^2 - \tau^2} \cdot e^{2i\rho\sigma\tau} \\ e^{2\sigma^2 - 2\tau^2 - 4i\rho\sigma\tau} - e^{\sigma^2 - \tau^2} \cdot e^{-2i\rho\sigma\tau} & e^{2\sigma^2} - e^{\sigma^2 - \tau^2} \end{bmatrix} = \begin{bmatrix} e^{2\sigma^2} & 0 \\ 0 & e^{2\sigma^2} \end{bmatrix}$$

Again, ϱ has disappeared from the limiting distribution; but in this case $\rho_{W\overline{W}} = 0$.

---

[12] For an event almost surely to happen means that its probability is unity; for an event almost surely not to happen means that its probability is zero. The latter case means not that the event will not happen, but rather that the event has zero probability mass. For example, if $X \sim$ Uniform[0, 1], Prob[$X=\frac{1}{2}$] = 0. So X almost surely does not equal ½, even though ½ is as possible as any other number in the interval.

[13] For more on this bimodal Arcsine(-1, 1) distribution see Wikipedia, "Arcsine distribution."

[14] The next section expands on this important subject. "Infinite variance in Y" means "as the variance of Y approaches infinity." It does not mean that e^{iY} is uniform for a variable Y whose variance is infinite, e.g., for a Pareto random variable whose shape parameter is less than or equal to two.

[15] Cf. Halliwell [2013] for a discussion on the right tails of the lognormal and other loss distributions.

In practical work with $U = \mathrm{Re}(W)$,[16] the angular part e^{iY} will be more important than the lognormal range e^X. For example, one who wanted the tendency for the larger magnitudes of $U = \mathrm{Re}(W)$ to be positive might set the mean of Y at $-\pi/2$ and the correlation ϱ to some positive value. Thus, greater than average values of Y, angling off into quadrants 4 and 1 of the complex plane, would correlate with larger than average values of X and hence of e^X. Of course, $Var[Y] = \tau^2$ would have to be small enough that deviations of $\pm\pi$ from $E[Y] = -\pi/2$ would be tolerably rare. Equivalently, one could set the mean of Y at $\pi/2$ and the correlation ϱ to some negative value. As a second example, one who wanted negative values of U to be less frequent than positive, might set both the mean of Y and ϱ to zero, and set the variance of Y so that $\mathrm{Prob}\big[|Y| > \pi/2\big]$ is desirably small. Some distributions of U for $\tau^2 \gg \sigma^2$ are bimodal, as in the specialized case $\sigma^2 \to 0^+$ and $\tau^2 \to \infty$. But less extreme parameters would result in unimodal distributions for U over the entire real number line.

10. THE COMPLEX UNIT-CIRCLE RANDOM VARIABLE

In the previous section we claimed that as the variance τ^2 of the normal random variable Y approaches infinity, e^{iY} approaches a uniform distribution over the complex unit circle. The explanation and justification of this claim in this section prepare for an important implication in the next.

Let real-valued random variable Y be distributed as $N[\mu, \sigma^2]$, and let $W = e^{iY}$. According to the moment-generating-formula of Section 8, $M_Y(it) = E[e^{itY}] = e^{it\mu + (it)^2\sigma^2/2} = e^{it\mu - t^2\sigma^2/2}$. Although the

---

[16] In the absence of an analytic distribution, practical work with the complex lognormal would seem to require simulating its values from the underlying normal distribution.

formula applies to complex values of t, here we'll restrict it to real values. With $t \in \Re$ $M_Y(it)$ is known as the characteristic function of real variable Y. And so:

$$\lim_{\sigma^2 \to \infty} M_Y(it) = \lim_{\sigma^2 \to \infty} e^{it\mu - t^2\sigma^2/2} = e^{it\mu} \lim_{\sigma^2 \to \infty} e^{-t^2\sigma^2/2} = \delta_{t0} = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases}$$

It is noteworthy, and indicative of a uniformity of some sort, that μ drops out of the result.

Next, let real-valued random variable Θ be uniformly distributed over $[a, a + 2\pi n]$, where n is a positive integer; in symbols, $\Theta \sim U[a, a + 2\pi n]$. Then:

$$M_\Theta(it) = E\left[e^{it\Theta}\right]$$

$$= \int_{\theta=a}^{a+2\pi n} e^{it\theta} \frac{1}{2\pi n} d\theta$$

$$= \left. \frac{e^{it\theta}}{2\pi itn} \right|_a^{a+2\pi n}$$

$$= e^{ita} \frac{e^{2\pi itn} - 1}{2\pi itn}$$

$$= \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } tn = \pm 1, \pm 2, \ldots \\ \neq 0 & \text{if } tn \text{ not integral} \end{cases}$$

Letting n approach infinity, we have:

$$\lim_{n \to \infty} M_\Theta(it) = e^{ita} \lim_{n \to \infty} \frac{e^{2\pi itn} - 1}{2\pi itn} = \delta_{t0} = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t = 0 \end{cases}$$

Hence, $\lim_{n \to \infty} M_\Theta(it) = \lim_{\sigma^2 \to \infty} M_Y(it) = \delta_{t0}$. The equality of the limits of the characteristic functions of the random variables implies the identity of the limits of their distributions; hence, the diffuse uniform $U[a, a + \infty]$ is "the same" as the diffuse normal $N[\mu, \infty]$.[17]

---

[17] Quotes are around 'the same' because the limiting distributions are not proper distributions. The notion of diffuse distributions comes from Venter [1996, pp. 406-410], who shows there how different diffuse distributions result in

Indeed, for the limit to be δ_{t0} it is not required that n be an integer. But for $\Theta \sim U[a, a + 2\pi n]$, the integral moments of $W = e^{i\Theta}$ are:

$$E[W^j] = E[e^{ij\Theta}] = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } jn = \pm 1, \pm 2, \ldots \\ \neq 0 & \text{if } jn \text{ not integral} \end{cases}$$

So if n is an integer, jn will be an integer, and all the integral moments of W will be zero, except for the zeroth. Therefore, the integral moments of $W = e^{i\Theta}$ are invariant to n, as long as the n in $2\pi n$, the width of the interval of Θ, is a whole number. Hence, although we hereby define the unit-circle random variable as $e^{i\Theta}$ for $\Theta \sim U[0, 2\pi]$, the choice of $a = 0$ and $n = 1$ is out of convenience, rather than out of necessity. The probability for $e^{i\Theta}$ to be in an arc of this circle of length l equals $l/2\pi$.

The integral moments of the conjugate of $\overline{W} = e^{-i\Theta}$ are the same, for $E[\overline{W}^j] = E[\overline{W^j}] = \overline{E[W^j]} = \overline{\delta_{j0}} = \delta_{j0} = E[W^j]$. Alternatively, $E[\overline{W}^j] = E[e^{-ij\Theta}] = \delta_{(-j)0} = \delta_{j0}$. And the jk^{th} mixed moment is $E[W^j \overline{W}^k] = E[e^{ij\Theta} e^{-ik\Theta}] = E[e^{i(j-k)\Theta}] = \delta_{(j-k)0} = \delta_{jk}$. Since $E[W] = E[\overline{W}] = 0$, the augmented variance of the unit-circle random variable is:

$$Var\begin{bmatrix} W \\ \overline{W} \end{bmatrix} = E\left[\begin{bmatrix} W \\ \overline{W} \end{bmatrix} \begin{bmatrix} \overline{W} & W \end{bmatrix} \right] = E\begin{bmatrix} W\overline{W} & WW \\ \overline{W}\,\overline{W} & \overline{W}W \end{bmatrix} = \begin{bmatrix} \delta_{11} & \delta_{20} \\ \delta_{02} & \delta_{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

Hence, $W = e^{i\Theta}$ for $\Theta \sim U[0, 2\pi]$ is not just a unit-circle random variable; having zero mean and unit variance, it is the *standard* unit-circle random variable.

---

different Bayesian estimates. But here every continuous random variable Y diffuses through the periodicity of e^{iY} into the same limiting distribution, viz., the Kronecker δ_{t0} (note 31).

Multiplying W by a complex constant $\alpha \neq 0$ affects the radius of the random variable, whose jk^{th} mixed moment is:

$$E\left[(\alpha W)^j \left(\overline{\alpha W}\right)^k\right] = \alpha^j \overline{\alpha}^k E\left[W^j \overline{W}^k\right] = \alpha^j \overline{\alpha}^k \delta_{jk} = \begin{cases} (\alpha\overline{\alpha})^j & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

The augmented variance is $Var\begin{bmatrix} \alpha W \\ \overline{\alpha W} \end{bmatrix} = \alpha\overline{\alpha} Var[W] = \alpha\overline{\alpha} \, I_2$. One may consider α as an instance of a complex random variable A. Due to the independence of A from W, the jk^{th} mixed moment of

AW is $E\left[(AW)^j \left(\overline{AW}\right)^k\right] = E\left[A^j \overline{A}^k\right] E\left[W^j \overline{W}^k\right] = E\left[A^j \overline{A}^k\right] \delta_{jk} = E\left[(A\overline{A})^j\right] \delta_{jk}$. Its augmented

variance is $Var\begin{bmatrix} AW \\ \overline{AW} \end{bmatrix} = E[A\overline{A}] Var[W] = \{Var[A] + E[A]E[\overline{A}]\} Var[W]$. Unlike the one-dimensional

W, AW can cover the whole complex plane. However, like W, it too possesses the desirable

property that $E\left[(AW)^j\right] = \delta_{j0}$.[18]

11. UNIT-CIRCULARITY AND DETERMINISM

The single most important quality of a random variable is its mean. In fact, just having reliable estimates of mean values would satisfy many users of actuarial analyses. Stochastic advances in actuarial science over the last few decades notwithstanding, much actuarial work remains deterministic. Determinism is not the reduction of a stochastic answer $Y = f(X)$ to its mean $E[Y] = E[f(X)]$. Rather, the deterministic assumption is that the expectation of a function of a random variable equals the function of the expectation of the random variable; in symbols,

---

[18] The existence of the moments $E\left[A^j \overline{A}^k\right]$ needs to be ascertained. In particular, moments for j and k as negative integers will not exist unless $\text{Prob}[A = 0] = 1 - \text{Prob}[A \neq 0] = 1 - \text{Prob}[A\overline{A} > 0] = 0$.

$E[Y] = E[f(X)] = f(E[X])$.   Because this assumption is true for linear f, it was felt to be a reasonable or necessary approximation for non-linear f.

Advances in computing hardware and software, as well as increased technical sophistication, have made determinism more avoidable and less acceptable. However, the complex unit-circular random variable provides a habitat for the survival of determinism. To see this, let f be analytic over the domain of complex random variable Z. From Cauchy's Integral Formula (Havil [2003, Appendix D.8 and D.9]) it follows that within the domain of Z, f can be expressed as a convergent series

$f(z) = a_0 + a_1 z + \ldots = a_0 + \sum_{j=1}^{\infty} a_j z^j$. Taking the expectation, we have:

$$E[f(Z)] = a_0 + \sum_{j=1}^{\infty} a_j E[Z^j]$$

But if for every positive integer j $E[Z^j] = E[Z]^j$, then:

$$E[f(Z)] = a_0 + \sum_{j=1}^{\infty} a_j E[Z^j] = a_0 + \sum_{j=1}^{\infty} a_j E[Z]^j = f(E[Z])$$

Therefore, determinism conveniently works for analytic functions of random variables whose moments are powers of their means.

Now a real-valued random variable whose moments are powers of its mean would have the characteristic function:

$$M_X(it) = E[e^{itX}] = 1 + \sum_{j=1}^{\infty} \frac{(it)^j}{j!} E[X^j] = 1 + \sum_{j=1}^{\infty} \frac{(it)^j}{j!} E[X]^j = e^{itE[X]} = M_{E[X]}(it)$$

This is the characteristic function of the "deterministic" random variable, i.e., the random variable whose probability is massed at one point, its mean. So determinism with real-valued random

variables requires "deterministic" random variables. But some complex random variables, such as the unit-circle, have the property $E\!\left[Z^{\,j}\right]= E[Z]^{j}$ without being deterministic.

In fact, when $E\!\left[Z^{\,j}\right]= E[Z]^{j}$, not only is $E[f(Z)]= f[E[Z]]$. For positive integer k, $f^{\,k}(z)$ is as analytic as f itself; hence, $E\!\left[f^{\,k}(Z)\right]= f^{\,k}(E[Z])$. So the determinism with these complex random variables is valid for all moments; nothing is lost.

In Section 10 we saw that for the unit-circle random variable $W = e^{i\Theta}$ and for $j = \pm 1, \pm 2, \ldots,$

$E\!\left[W^{-j}\right]= E\!\left[W^{\,j}\right]= 0 = E[W]^{|j|}$. Can determinism extend to non-analytic functions which involve the negative moments? For example, let $g(z)= 1/(\eta - z)$, for some complex $\eta \neq 0$. The function is singular at $z = \eta$; but within the disc $\{z : |z/\eta| < 1\}= \{z : |z| < |\eta|\}$ the function equals the convergent series:

$$g(z)= 1/(\eta - z)= \frac{1}{\eta}\cdot\frac{1}{\left(1-\dfrac{z}{\eta}\right)}= \frac{1}{\eta}\left\{1+\frac{z}{\eta}+\left(\frac{z}{\eta}\right)^{2}+\ldots\right\}= \frac{1}{\eta}+\frac{z}{\eta^{2}}+\frac{z^{2}}{\eta^{3}}+\ldots$$

Outside the disc, or for $\{z : |z/\eta| > 1\}= \{z : |z| > |\eta|\}$, another convergent series represents the function:

$$g(z)= 1/(\eta - z)= -\frac{1}{z}\cdot\frac{1}{\left(1-\dfrac{\eta}{z}\right)}= -\frac{1}{z}\left\{1+\frac{\eta}{z}+\left(\frac{\eta}{z}\right)^{2}+\ldots\right\}= -\frac{1}{z}-\frac{\eta}{z^{2}}-\frac{\eta^{2}}{z^{3}}-\ldots$$

So, if $|\eta| > 1$, then $|W| = 1 < |\eta|$. In this case:

$$E[g(W)] = E\left[\frac{1}{\eta} + \frac{W}{\eta^2} + \frac{W^2}{\eta^3} + \ldots\right]$$

$$= \frac{1}{\eta} + \frac{E[W]}{\eta^2} + \frac{E[W^2]}{\eta^3} + \ldots$$

$$= \frac{1}{\eta} + \frac{0}{\eta^2} + \frac{0}{\eta^3} + \ldots$$

$$= \frac{1}{\eta}$$

However, if $|\eta| < 1$, then $|W| = 1 > |\eta|$. So in this case:

$$E[g(W)] = E\left[-\frac{1}{W} - \frac{\eta}{W^2} - \frac{\eta^2}{W^3} - \ldots\right]$$

$$= -E[W^{-1}] - \eta E[W^{-2}] - \eta^2 E[W^{-3}] - \ldots$$

$$= -0 - \eta \cdot 0 - \eta^2 \cdot 0 - \ldots$$

$$= 0$$

Both answers are correct; however, only the first satisfies the deterministic equation

$$E[g(W)] = g(E[W]) = g(0) = 1/\eta.$$

To understand why the answer depends on whether η is inside or outside the complex unit circle, let us evaluate $E[g(W)]$ directly:

$$E[g(W)] = E[1/(\eta - W)] = E\left[1/\left(\eta - e^{i\Theta}\right)\right] = \int_{\theta=0}^{2\pi} \frac{1}{\eta - e^{i\theta}} \frac{d\theta}{2\pi}$$

The next step is to transform from θ into $z = e^{i\theta}$. So $dz = ie^{i\theta} d\theta = iz d\theta$, and the line integral transforms into a contour integral over the unit circle C:

$$E[g(W)] = \int_{\theta=0}^{2\pi} \frac{1}{\eta - e^{i\theta}} \frac{d\theta}{2\pi}$$

$$= \oint_C \frac{1}{\eta - z} \frac{iz}{iz} \frac{d\theta}{2\pi}$$

$$= \oint_C \frac{1}{\eta - z} \frac{1}{z} \frac{dz}{2\pi i}$$

$$= \frac{1}{2\pi i} \oint_C \frac{dz}{z(\eta - z)}$$

$$= \frac{1}{2\pi i} \oint_C \left(\frac{1}{z} + \frac{1}{\eta - z} \right) \frac{1}{\eta} dz$$

$$= \frac{1}{\eta} \left(\frac{1}{2\pi i} \oint_C \frac{dz}{z} + \frac{1}{2\pi i} \oint_C \frac{dz}{\eta - z} \right)$$

$$= \frac{1}{\eta} \left(\frac{1}{2\pi i} \oint_C \frac{dz}{z-0} - \frac{1}{2\pi i} \oint_C \frac{dz}{z-\eta} \right)$$

Now the value of each of these integrals is one if its singularity is within the unit circle C, and zero if it is not.[19] Of course, the singularity of the first integral at $z = 0$ lies within C; hence, its value is one. The second integral's singularity at $z = \eta$ lies within C if and only if $|\eta| < 1$. Therefore:

$$E[g(W)] = \frac{1}{\eta} \left(\frac{1}{2\pi i} \oint_C \frac{dz}{z-0} - \frac{1}{2\pi i} \oint_C \frac{dz}{z-\eta} \right) = g(E[W]) \cdot \begin{cases} 1 & \text{if } |\eta| > 1 \\ 0 & \text{if } |\eta| < 1 \end{cases}$$

So the deterministic equation will hold for one of the Laurent series according to which the domain of the non-analytic function is divided into regions of convergence. Fascinating enough is how the

function $\varphi(\eta) = \frac{1}{2\pi i} \oint_C \frac{dz}{z-\eta} = \begin{cases} 1 & \text{if } |\eta| > 1 \\ 0 & \text{if } |\eta| < 1 \end{cases}$ serves as the indicator of a state, viz., the state of being

inside or outside the complex unit circle.

---

[19] Technically, the integral has no value if the singularity lies on C; but there are some practical advantages for "splitting the difference" in that case.

12. THE LINEAR STATISTICAL MODEL

Better known as "regression" models, linear statistical models extend readily into the realm of complex numbers. A general real-valued form of such models is presented and derived in Halliwell [1997, Appendix C]:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}, \quad Var\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

The subscript '1' denotes observations, '2' denotes predictions. Vector $\mathbf{y}_1$ is observed; the whole design matrix X is hypothesized, as well as the fourfold 'Σ' variance structure. Although the variance structure may be non-negative-definite (NND), the variance of the observations Σ_{11} must be positive-definite (PD). Also, the observation design X_1 must be of full column rank. The last two requirements ensure the existence of the inverses Σ_{11}^{-1} and $\left(X_1'\Sigma_{11}^{-1}X_1\right)^{-1}$. The best linear unbiased predictor of $\mathbf{y}_2$ is $\hat{\mathbf{y}}_2 = X_2\hat{\beta} + \Sigma_{21}\Sigma_{11}^{-1}\left(\mathbf{y}_1 - X_1\hat{\beta}\right)$. The variance of prediction error $\mathbf{y}_2 - \hat{\mathbf{y}}_2$ is $Var\left[\mathbf{y}_2 - \hat{\mathbf{y}}_2\right] = \left(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1\right)Var\left[\hat{\beta}\right]\left(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1\right)' + \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Embedded in these formulas are the estimator of β and its variance:

$$\hat{\beta} = \left(X_1'\Sigma_{11}^{-1}X_1\right)^{-1}X_1'\Sigma_{11}^{-1}\mathbf{y}_1 = Var\left[\hat{\beta}\right] \cdot X_1'\Sigma_{11}^{-1}\mathbf{y}_1.$$

For the purpose of introducing complex numbers into the linear statistical model we will concern ourselves here only the estimation of the parameter β. So we drop the subscripts '1' and '2' and simplify the observation as $\mathbf{y} = X\beta + \mathbf{e}$, where $Var[\mathbf{e}] = \Gamma$. Again, X must be of full column rank and Γ must be Hermetian PD. According to Section 4, transjugation is to complex matrices what transposition is to real-valued matrices. Therefore, the short answer for a complex model is:

$$\hat{\beta} = \left(X^*\Gamma^{-1}X\right)^{-1}X^*\Gamma^{-1}\mathbf{y} = Var\left[\hat{\beta}\right] \cdot X^*\Gamma^{-1}\mathbf{y}.$$

However, deriving the solution from the double-real representation in Section 3 will deepen the understanding. The double-real form of the observation is:

$$\begin{bmatrix} \mathbf{y}_r & -\mathbf{y}_i \\ \mathbf{y}_i & \mathbf{y}_r \end{bmatrix} = \begin{bmatrix} X_r & -X_i \\ X_i & X_r \end{bmatrix} \begin{bmatrix} \beta_r & -\beta_i \\ \beta_i & \beta_r \end{bmatrix} + \begin{bmatrix} \mathbf{e}_r & -\mathbf{e}_i \\ \mathbf{e}_i & \mathbf{e}_r \end{bmatrix}$$

All the vectors and matrices in this form are real-valued. The subscripts '*r*' and '*i*' denote the real and imaginary parts of $\mathbf{y}$, X, β, and $\mathbf{e}$. Due to the redundancy of double-real representation, we may retain just the left column:

$$\begin{bmatrix} \mathbf{y}_r \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} X_r & -X_i \\ X_i & X_r \end{bmatrix} \begin{bmatrix} \beta_r \\ \beta_i \end{bmatrix} + \begin{bmatrix} \mathbf{e}_r \\ \mathbf{e}_i \end{bmatrix}$$

Note that if X is real-valued, then $X_i = 0$, and $\mathbf{y}_r$ and $\mathbf{y}_i$ become two "data panels," each with its own parameter β_r and β_i.[20]

Now let $\Xi_t = \begin{bmatrix} I_t & iI_t \\ I_t & -iI_t \end{bmatrix}$, the augmentation matrix of Section 7, where *t* is the number of observations. Since the augmentation matrix is non-singular, premultiplying the left-column form by it yields equivalent but insightful forms:

---

[20] This assumes zero covariance between the error vectors.

$$\begin{bmatrix} I_t & iI_t \\ I_t & -iI_t \end{bmatrix}\begin{bmatrix} \mathbf{y}_r \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} I_t & iI_t \\ I_t & -iI_t \end{bmatrix}\begin{bmatrix} X_r & -X_i \\ X_i & X_r \end{bmatrix}\begin{bmatrix} \beta_r \\ \beta_i \end{bmatrix} + \begin{bmatrix} I_t & iI_t \\ I_t & -iI_t \end{bmatrix}\begin{bmatrix} \mathbf{e}_r \\ \mathbf{e}_i \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y}_r + i\mathbf{y}_i \\ \mathbf{y}_r - i\mathbf{y}_i \end{bmatrix} = \begin{bmatrix} X_r + iX_i & -X_i + iX_r \\ X_r - iX_i & -X_i - iX_r \end{bmatrix}\begin{bmatrix} \beta_r \\ \beta_i \end{bmatrix} + \begin{bmatrix} \mathbf{e}_r + i\mathbf{e}_i \\ \mathbf{e}_r - i\mathbf{e}_i \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X & iX \\ \overline{X} & -i\overline{X} \end{bmatrix}\begin{bmatrix} \beta_r \\ \beta_i \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X\beta_r + iX\beta_i \\ \overline{X}\beta_r - i\overline{X}\beta_i \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X\beta \\ \overline{X}\overline{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}\begin{bmatrix} \beta \\ \overline{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$$

The first insight is that $\overline{\mathbf{y}} = \overline{X\beta + \mathbf{e}} = \overline{X}\overline{\beta} + \overline{\mathbf{e}}$ is as much observed as $\mathbf{y} = X\beta + \mathbf{e}$. The second

insight is that $Var\begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$ is an augmented variance, whose general form according to Section 6 is

$Var\begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}$. Therefore, the general form of the observation of a complex linear model is

$\begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}\begin{bmatrix} \beta \\ \overline{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$, where $Var\begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}$. Not only is $\overline{\mathbf{y}}$ as observable as $\mathbf{y}$, but also $\overline{\beta}$

is as estimable as β. Furthermore, although the augmented variance may default to $C = 0$, the

complex linear statistical model does not require $\begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$ to be "proper complex," as defined in Section

7.

Since X is of full column rank, so too must be $\begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}$. And since Γ is Hermetian PD, both it and

its conjugate $\overline{\Gamma}$ are invertible. But the general form of the observation requires $\begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}$ to be

Hermetian PD, hence invertible. A consequence is that both the "determinant" forms $\Gamma - C\overline{\Gamma}^{-1}\overline{C}$

and $\overline{\Gamma} - \overline{C}\Gamma^{-1}C$ are Hermetian PD and invertible. With this background it can be shown, and the

reader should verify, that $\begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}^{-1} = \begin{bmatrix} H & K \\ \overline{K} & \overline{H} \end{bmatrix}$, where $H = (\Gamma - C\overline{\Gamma}^{-1}\overline{C})^{-1}$ and $K = -\Gamma^{-1}C\overline{H}$. The

important point is that inversion preserves the augmented-variance form.

The solution of the complex linear model $\begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}\begin{bmatrix} \beta \\ \overline{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix}$, where $Var\begin{bmatrix} \mathbf{e} \\ \overline{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}$, is:

$$
\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\overline{\boldsymbol{\beta}}} \end{bmatrix} = \left(\begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}^* \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}^{-1} \begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix} \right)^{-1} \begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}^* \begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix}
$$

$$
= \left(\begin{bmatrix} X^* & 0 \\ 0 & X' \end{bmatrix} \begin{bmatrix} H & K \\ \overline{K} & \overline{H} \end{bmatrix} \begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix} \right)^{-1} \begin{bmatrix} X^* & 0 \\ 0 & X' \end{bmatrix} \begin{bmatrix} H & K \\ \overline{K} & \overline{H} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \overline{\mathbf{y}} \end{bmatrix}
$$

$$
= \begin{bmatrix} X^*HX & X^*K\overline{X} \\ X'\overline{K}X & X'\overline{H}\overline{X} \end{bmatrix}^{-1} \begin{bmatrix} X^*H\mathbf{y} + X^*K\overline{\mathbf{y}} \\ X'\overline{K}\mathbf{y} + X'\overline{H}\overline{\mathbf{y}} \end{bmatrix}
$$

And $Var\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\overline{\boldsymbol{\beta}}} \end{bmatrix} = \begin{bmatrix} X^*HX & X^*K\overline{X} \\ X'\overline{K}X & X'\overline{H}\overline{X} \end{bmatrix}^{-1}$, which must exist since it is a quadratic form based on the

Hermetian PD $\begin{bmatrix} \Gamma & C \\ \overline{C} & \overline{\Gamma} \end{bmatrix}^{-1}$ and the full column rank $\begin{bmatrix} X & 0 \\ 0 & \overline{X} \end{bmatrix}$. Reformulate this as:

$$
\begin{bmatrix} X^*HX & X^*K\overline{X} \\ X'\overline{K}X & X'\overline{H}\overline{X} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\overline{\boldsymbol{\beta}}} \end{bmatrix} = \begin{bmatrix} X^*H\mathbf{y} + X^*K\overline{\mathbf{y}} \\ X'\overline{K}\mathbf{y} + X'\overline{H}\overline{\mathbf{y}} \end{bmatrix}
$$

The conjugates of the two equations in $\hat{\boldsymbol{\beta}}$ and $\hat{\overline{\boldsymbol{\beta}}}$ are the same equations in $\hat{\overline{\boldsymbol{\beta}}}$ and $\hat{\boldsymbol{\beta}}$:

$$
\begin{bmatrix} X^*HX & X^*K\overline{X} \\ X'\overline{K}X & X'\overline{H}\overline{X} \end{bmatrix} \begin{bmatrix} \hat{\overline{\boldsymbol{\beta}}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} X^*H\mathbf{y} + X^*K\overline{\mathbf{y}} \\ X'\overline{K}\mathbf{y} + X'\overline{H}\overline{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} X^*HX & X^*K\overline{X} \\ X'\overline{K}X & X'\overline{H}\overline{X} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\overline{\boldsymbol{\beta}}} \end{bmatrix}
$$

Therefore, $\begin{bmatrix} \overline{\hat{\boldsymbol{\beta}}} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\overline{\boldsymbol{\beta}}} \end{bmatrix}$. It is well known that the estimator of a linear function of a random variable

is the linear function of the estimator of the random variable. But conjugation is not a linear function. Nevertheless, we have just proven that the estimator of the conjugate is the conjugate of the estimator.

13. ACTUARIAL APPLICATIONS OF COMPLEX RANDOM VARIABLES

How might casualty actuaries put complex random variables to work? Since the support of most complex random variables is a plane, rather than a line, their obvious application is bivariate. An example is a random variable whose real part is loss and whose imaginary part is LAE. Another application might pertain to copulas. According to Venter [2002, p. 69], "copulas are joint distributions of unit random variables." One could translate these joint distributions into distributions of complex variables whose support is the complex unit square, i.e., the square whose vertices are the points $z = 0, 1, 1+i, i$. However, for now it seems that real-valued bivariates provide the necessary theory and technique for these purposes.

Actuaries who have applied log-linear models to triangles with paid increments have been frustrated applying them to incurred triangles. The problem is that incurred increments are often negative, and the logarithm of a negative number is not real-valued. This has led Glenn Meyers [2013] to seek modified lognormal distributions whose support includes the negative real numbers. The persistent intractability of the log-linear problem was a major reason for our attention to the lognormal random vector $\mathbf{w}_{n \times 1} = e^{\mathbf{z}_{n \times 1}}$ in Section 8. But to model an incurred loss as the exponential function of a complex number suffers from two drawbacks. First, to model a real-valued loss as $e^{\mathbf{z}} = e^x \cdot e^{iy}$ requires y to be an integral multiple of π. The mixed random variable e^X with probability p and

$-e^X$ with probability $1-p$ is not lognormal. No more suitable are such "denatured" random variables as $\text{Re}(e^Z)$. Second, one still cannot model the eminently practical value of zero, because for all z, $e^z \neq 0$.[21] At present it does not appear that complex random variables will give birth to useful distributions of real-valued random variables. Even the unit-circle and indicator random variables of Sections 10 and 11, as interesting as they are in the theory of analytic functions, most likely will engender no distributions valuable to actuarial work.

The complex version of the linear model in Section 12 showed us that conjugates of observations are themselves observations and that conjugates of estimators are estimators of conjugates. Moreover, there we found a use for augmented variance. Nonetheless we are still fairly bound to our conclusion to Section 3, that one who lacked either the confidence or the software to work with complex numbers could probably do a work-around with double-real matrices.

So how can actuarial science benefit from complex random variables? The great benefit will come from new ways of thinking. The first step will be to overcome the habit of picturing a complex number as half real and half imaginary. Historically, it was only after numbers had expanded from rational to irrational that the whole set was called "real." Numbers ultimately are sets; zero is just the empty set. How real are sets? Regardless of their mathematical reality, they are not physically real. Complex numbers were deemed "real" because mathematicians needed them for the solution of polynomial equations. In the nineteenth century this spurred the development of abstract algebra. At first new ways of thinking amount to differences in degree; at some point many develop

---

[21] If $e^a = 0$ for some a, then for all z $e^z = e^{z-a+a} = e^{z-a}e^a = e^{z-a} \cdot 0 = 0$. One who sees that $\lim_{x \to -\infty} e^x \cdot e^{iy} = 0 \cdot e^{iy} = 0$ might propose to add the ordinate $\text{Re}(z) = x = -\infty$ to the complex plane. But not only is this proposal artificial; it also militates against the standard theory of complex variables, according to which all points infinitely far from zero constitute one and the same point at infinity.

into differences in kind. One might argue, "Why study Euclidean geometry? It all derives from a few axioms." True, but great theorems (e.g., that the sum of the angles of a triangle is the sum of two right angles) can be a long way from their axioms. A theorem means more than the course of its proof; often there are many proofs of a theorem. Furthermore, mathematicians often work backwards from accepted or desired truths to efficient and elegant sets of axioms. Perhaps the most wonderful thing about mathematics is its "unreasonable effectiveness in the natural sciences," to quote physicist Eugene Wigner. The causality between pure and applied mathematics works in both directions. Therefore, it is likely that complex random variables and vectors will find their way into actuarial science. But it will take years, even decades, and technology and education will have to prepare for it.

14. CONCLUSION

Just as physics divides into different areas, e.g., theoretical, experimental, and applied, so too actuarial science, though perhaps more concentrated on business application, justifiably has and needs a theoretical component. Theory and application cross-fertilize each other. In this paper we have proposed to add complex numbers to the probability and statistics of actuarial theory. With patience, the technically inclined actuary should be able to understand the theory of complex random variables delineated herein. In fact, our multivariate approach may even more difficult to understand than the complex-function theory; but both belong together. Although all complex matrices and operations were formed from double-real counterparts, we believe that the "sum is greater than the parts," i.e., that the assimilation of this theory will lead to higher-order thinking and creativity. In the sixteenth century the "fiction" of $i = \sqrt{-1}$ allowed mathematicians to solve more equations. Although at first complex solutions were deemed "extraneous roots," eventually their practicality became recognized; so that today complex numbers are essential for science and

engineering. Applying complex numbers to probability has lagged; but even now it is part of signal processing in electrical engineering. Knowing how rapidly science has developed with nuclear physics, molecular biology, space exploration, and computers, who would dare to bet against the usefulness of complex random variables to actuarial science by the mid-2030s, when many scientists and futurists expect nuclear fusion to be harnessed and available for commercial purposes?

REFER**ENCES**

[1.] Halliwell, Leigh J., Conjoint Prediction of Paid and Incurred Losses, CAS *Forum*, Summer 1997, 241-380, www.casact.org/pubs/forum/97sforum/97sf1241.pdf.

[2.] Halliwell, Leigh J., "Classifying the Tails of Loss Distributions," CAS *E-Forum*, Spring 2013, Volume 2, www.casact.org/pubs/forum/13spforumv2/Haliwell.pdf.

[3.] Havil, Julian, *Gamma: Exploring Euler's Constant*, Princeton University Press, 2003.

[4.] Healy, M. J. R., *Matrices for Statistics*, Oxford, Clarendon Press, 1986.

[5.] Hogg, Robert V., and Stuart A. Klugman, *Loss Distributions*, New York, Wiley, 1984.

[6.] Johnson, Richard A., and Dean Wichern, *Applied Multivariate Statistical Analysis* (Third Edition), Englewood Cliffs, NJ, Prentice Hall, 1992.

[7.] Judge, George G., Hill, R. C., et al., *Introduction to the Theory and Practice of Econometrics* (Second Edition), New York, Wiley, 1988.

[8.] Klugman, Stuart A., et al., *Loss Models: From Data to Decisions*, New York, Wiley, 1998.

[9.] Meyers, Glenn, "The Skew Normal Distribution and Beyond," *Actuarial Review*, May 2013, p. 15, www.casact.org/newsletter/pdfUpload/ar/AR_May2013_1.pdf.

[10.] Million, Elizabeth, The Hadamard Product, 2007, http://buzzard.ups.edu/courses/2007spring/projects/million-paper.pdf.

[11.] Schneider, Hans, and George Barker, *Matrices and Linear Algebra*, New York, Dover, 1973.

[12.] Veeravalli, V. V., *Proper Complex Random Variables and Vectors*, 2006, http://courses.engr.illinois.edu/ece461/handouts/notes6.pdf.

[13.] Venter, G.G., "Credibility," *Foundations of Casualty Actuarial Science* (Third Edition), Casualty Actuarial Society, 1996.

[14.] Venter, G.G., "Tails of Copulas," *PCAS* LXXXIX, 2002, 68-113, www.casact.org/pubs/proceed/proceed02/02068.pdf.

[15.] Weyl, Hermann, *The Theory of Groups and Quantum Mechanics*, New York, Dover, 1950 (ET by H. P. Robertson from second German Edition 1930).

[16.] Wikipedia contributors, "Arcsine distribution," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Arcsine_distribution (accessed October 2014).

[17.] Wikipedia contributors, "Complex normal distribution," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Complex_normal_distribution (accessed October 2014).

[18.] Wikipedia contributors, "Complex number," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Complex_number (accessed October 2014).

APPENDIX A

REAL-VALUED LOGNORMAL RANDOM VECTORS

Feeling that the treatment of lognormal random vectors in Section 8 would be too long, we have decided to prepare for it in Appendices A and B. According to Section 7, the probability density function of real-valued $n \times 1$ normal random vector $\mathbf{x}$ with mean μ and variance Σ is:

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}$$

Therefore, $\displaystyle\int_{x \in \mathfrak{R}^n} f_{\mathbf{x}}(\mathbf{x}) dV = 1$. The single integral over $\mathfrak{R}^n$ represents an n-multiple integral over each

x_j from $-\infty$ to $+\infty$; $dV = dx_1 \ldots dx_n$.

The moment generating function of $\mathbf{x}$ is $M_{\mathbf{x}}(t) = E[e^{t'\mathbf{x}}] = E\left[e^{\sum_{j=1}^{n} t_j x_j}\right]$, where t is a suitable $n \times 1$

vector.[22] Partial derivatives of the moment generating function evaluated at $t = 0_{n \times 1}$ equal moments of $\mathbf{x}$, since:

$$\left.\frac{\partial^{k_1 + \ldots k_n} M_{\mathbf{x}}(t)}{\partial^{k_1} x_1 \ldots \partial^{k_n} x_n}\right|_{t=0} = E\left[x_1^{k_1} \ldots x_n^{k_n} e^{t'\mathbf{x}}\right]\Big|_{t=0} = E\left[x_1^{k_1} \ldots x_n^{k_n}\right]$$

But lognormal moments are values of the function itself. For example, if $t = e_j$, the jth unit vector,

then $M_{\mathbf{x}}(e_j) = E[e^{e_j'\mathbf{x}}] = E[e^{X_j}]$. Likewise, $M_{\mathbf{x}}(e_j + e_k) = E[e^{X_j} e^{X_k}]$. The moment generating function of $\mathbf{x}$, if it exists, is the key to the moments of $e^{\mathbf{x}}$.

---

[22] All real-valued t vectors are suitable; Appendix B will extend the suitability to complex t.

The moment generating function of the real-valued multivariate normal **x** is:

$$M_{\mathbf{x}}(t) = E\left[e^{t'\mathbf{x}}\right]$$

$$= \int_{x \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} e^{t'\mathbf{x}} dV$$

$$= \int_{x \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}\left\{(x-\mu)'\Sigma^{-1}(x-\mu)-2t'\mathbf{x}\right\}} dV$$

A multivariate "completion of the square" results in the identity:

$$(x-\mu)'\Sigma^{-1}(x-\mu) - 2t'x = \left(x-[\mu+\Sigma t]\right)'\Sigma^{-1}\left(x-[\mu+\Sigma t]\right) - 2t'\mu - t'\Sigma t$$

We leave it for the reader to verify. By substitution, we have:

$$M_{\mathbf{x}}(t) = \int_{x \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}\left\{(x-\mu)'\Sigma^{-1}(x-\mu)-2t'x\right\}} dV$$

$$= \int_{x \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}\left\{(x-[\mu+\Sigma t])'\Sigma^{-1}(x-[\mu+\Sigma t])-2t'\mu-t'\Sigma t\right\}} dV$$

$$= \int_{x \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-[\mu+\Sigma t])'\Sigma^{-1}(x-[\mu+\Sigma t])} dV \cdot e^{t'\mu + t'\Sigma t/2}$$

$$= 1 \cdot e^{t'\mu + t'\Sigma t/2}$$

$$= e^{t'\mu + t'\Sigma t/2}$$

The reduction of the integral to unity in the second last line is due to the fact that

$\dfrac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-[\mu+\Sigma t])'\Sigma^{-1}(x-[\mu+\Sigma t])}$ is the probability density function of the real-valued $n\times 1$ normal

random vector with mean $\mu + \Sigma t$ and variance Σ. This new mean is valid if it is real-valued, which

will be so if t is real-valued. In fact, $\mu + \Sigma t$ is real-valued if and only if t is real-valued.

So the moment generating function of the real-valued normal multivariate $\mathbf{x} \sim N(\mu, \Sigma)$ is $M_{\mathbf{x}}(t) = e^{t'\mu + t'\Sigma t/2}$, which is valid at least for $t \in \mathfrak{R}^n$. As a check:[23]

$$\frac{\partial M_{\mathbf{x}}(t)}{\partial t} = (\mu + \Sigma t)e^{t'\mu + t'\Sigma t/2} \Rightarrow E[\mathbf{x}] = \frac{\partial M_{\mathbf{x}}(t)}{\partial t}\bigg|_{t=0} = \mu$$

And for the second derivative:

$$\frac{\partial^2 M_{\mathbf{x}}(t)}{\partial t \partial t'} = \frac{\partial(\mu + \Sigma t)e^{t'\mu + t'\Sigma t/2}}{\partial t'}$$

$$= \left(\Sigma + (\mu + \Sigma t)(\mu + \Sigma t)' \right)e^{t'\mu + t'\Sigma t/2}$$

$$\Rightarrow E[\mathbf{x}\mathbf{x}'] = \frac{\partial^2 M_{\mathbf{x}}(t)}{\partial t \partial t'}\bigg|_{t=0} = \Sigma + \mu\mu'$$

$$\Rightarrow Var[\mathbf{x}] = E[\mathbf{x}\mathbf{x}'] - \mu\mu' = \Sigma$$

The lognormal moments follow from the moment generating function:

$$E[e^{X_j}] = E[e^{e'_j\mathbf{x}}] = M_{\mathbf{x}}(e_j) = e^{e'_j\mu + e'_j\Sigma e_j/2} = e^{\mu_j + \Sigma_{jj}/2}$$

The second moments are conveniently expressed in terms of first moments:

$$E[e^{X_j}e^{X_k}] = E\left[e^{(e_j+e_k)'\mathbf{x}}\right]$$

$$= e^{(e_j+e_k)'\mu + (e_j+e_k)'\Sigma(e_j+e_k)/2}$$

$$= e^{\mu_j + \mu_k + (\Sigma_{jj} + \Sigma_{jk} + \Sigma_{kj} + \Sigma_{kk})/2}$$

$$= e^{\mu_j + \Sigma_{jj}/2 + \mu_k + \Sigma_{kk}/2 + (\Sigma_{jk} + \Sigma_{kj})/2}$$

$$= e^{\mu_j + \Sigma_{jj}/2} \cdot e^{\mu_k + \Sigma_{kk}/2} \cdot e^{(\Sigma_{jk} + \Sigma_{kj})/2}$$

$$= e^{\mu_j + \Sigma_{jj}/2} \cdot e^{\mu_k + \Sigma_{kk}/2} \cdot e^{(\Sigma_{jk} + \Sigma_{jk})/2}$$

$$= E[e^{X_j}]E[e^{X_k}] \cdot e^{\Sigma_{jk}}$$

---

[23] The vector formulation of partial differentiation is explained in Appendix A.17 of Judge [1988].

So, $Cov[e^{X_j}, e^{X_k}] = E[e^{X_j} e^{X_k}] - E[e^{X_j}] E[e^{X_k}] = E[e^{X_j}] E[e^{X_k}] (e^{\Sigma_{jk}} - 1)$, which is the multivariate

equivalent of the well-known scalar formula $CV^2[e^X] = Var[e^X] / E[e^X]^2 = e^{\sigma^2} - 1$. Letting $E[e^{\mathbf{x}}]$

denote the $n \times 1$ vector whose j^{th} element is $E[e^{X_j}]$,[24] and $diag\left(E[e^{\mathbf{x}}] \right)$ as its $n \times n$ diagonalization,

we have $Var[\mathbf{x}] = diag\left(E[e^{\mathbf{x}}] \right) \{ e^{\Sigma} - 1_{n \times n} \} diag\left(E[e^{\mathbf{x}}] \right)$. Because $diag\left(E[e^{\mathbf{x}}] \right)$ is diagonal in positive

elements (hence, symmetric and PD), $Var[\mathbf{x}]$ is NND [or PD] if and only if $e^{\Sigma} - 1_{n \times n}$ is NND [or

PD].[25] Symmetry is no issue here, because for real-valued matrices, Σ is symmetric if and only if

$e^{\Sigma} - 1_{n \times n}$ is symmetric.

The relation between Σ and $T = e^{\Sigma} - 1_{n \times n}$ is merits a discussion whose result will be clear from a

consideration of 2×2 matrices. Since $\Sigma_{2 \times 2}$ is symmetric, it is defined in terms of three real numbers:

$\Sigma = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$. Now Σ is NND if and only if 1) $a \geq 0$, 2) $d \geq 0$, and 3) $ad - b^2 \geq 0$. Σ is PD if and

only if these three conditions are strictly greater than zero. If a or d is zero, by the third condition b

also must be zero.[26] Since we are not interested in degenerate random variables, which are

effectively constants, we will require a and d to be positive. With this requirement, Σ is NND if and

---

[24] This would follow naturally from the "elementwise" interpretation of $e^{\mathbf{A}}$, i.e., that the exponential function of matrix A is the matrix of the exponential functions of the elements of A. But if A is a square matrix, $e^{\mathbf{A}}$ may have the "matrix" interpretation $\mathbf{I}_n + \sum_{j=1}^{\infty} \mathbf{A}^j / j!$.

[25] PD [positive-definite] and NND [non-negative-definite] are defined in Section 4.

[26] Since NND matrices represent variances, $\Sigma = Var\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] \\ Cov[X_2, X_1] & Cov[X_2, X_2] \end{bmatrix} = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$. The fact that a or d equals 0 implies that b equals 0 means that a random variable can't covary with another random variable unless it covaries with itself.

only if $b^2 \leq ad$, and PD if and only if $b^2 < ad$. Since a and d are positive, so too is ad, as well as

the geometric mean $\gamma = \sqrt{ad}$. So Σ is NND if and only if $-\gamma \leq b \leq \gamma$, and PD if and only if

$-\gamma < b < \gamma$. It is well-known that $\min(a, d) \leq \gamma \leq \dfrac{a+d}{2} \leq \max(a, d)$ with equality if and only if

$a = d$.

Now the same three conditions determine the definiteness of $T = e^{\Sigma} - 1_{2\times2} = \begin{bmatrix} e^a - 1 & e^b - 1 \\ e^b - 1 & e^d - 1 \end{bmatrix}$.

Since we required a and d to be positive, both $e^a - 1$ and $e^d - 1$ are positive. This leaves the

definiteness of T dependent on the relation between $(e^b - 1)(e^b - 1)$ and $(e^a - 1)(e^d - 1)$. We will

next examine this relation according to the three cases $b = 0$, $b > 0$, and $b < 0$, all of which must

be subject to $-\gamma \leq b \leq \gamma$.

First, if $b = 0$, then $-\gamma < b < \gamma$ and Σ is PD. Furthermore, $(e^b - 1)(e^b - 1) = 0 < (e^a - 1)(e^d - 1)$.

Therefore, in this case, the lognormal transformation $\Sigma \to T = e^{\Sigma} - 1_{2\times2}$ is from PD to PD. And

zero covariance in the normal pair produces zero covariance in the lognormal pair. In fact, since

zero covariance between normal bivariates implies independence (cf. §2.5.7 of Judge [1988]), the

lognormal bivariates also are independent.

In the second case, $b > 0$, or more fully, $0 < b \leq \gamma$. Σ is PD if and only if $b < \gamma$. Define the

function $\varphi(x) = \ln\big((e^x - 1)/x\big)$ for positive real x (or $x \in \Re^+$). A graph will show that the function

strictly increases, i.e., $\varphi(x_1) < \varphi(x_2)$ if and only if $x_1 < x_2$. Moreover, the function is concave

upward. This means that the line segment between points $(x_1, \varphi(x_1))$ and $(x_2, \varphi(x_2))$ lies above

the curve $\varphi(x)$ for intermediate values of x. In particular, $\dfrac{\varphi(x_1) + \varphi(x_2)}{2} \geq \varphi\left(\dfrac{x_1 + x_1}{2}\right)$.

Equivalently, for all x_1, $x_2 \in \Re^+$, $2\varphi\left(\dfrac{x_1 + x_1}{2}\right) \leq \varphi(x_1) + \varphi(x_2)$ with equality if and only if $x_1 = x_2$.

Therefore, since a and d are positive, $2\varphi\left(\dfrac{a+d}{2}\right) \leq \varphi(a) + \varphi(d)$. And since $0 < \gamma = \sqrt{ad} \leq \dfrac{a+d}{2}$,

$2\varphi(\gamma) \leq 2\varphi\left(\dfrac{a+d}{2}\right) \leq \varphi(a) + \varphi(d)$. So $2\varphi(\gamma) \leq \varphi(a) + \varphi(d)$ with equality if and only if $a = d$.

Furthermore, since in this case $0 < b \leq \gamma$, $2\varphi(b) \leq 2\varphi(\gamma) \leq \varphi(a) + \varphi(d)$. Hence,

$2\varphi(b) \leq \varphi(a) + \varphi(d)$. The equality prevails if and only if $b = \gamma = a = d$, or if and only if $a = b = d$.

If $a = b = d$ then Σ is not PD; otherwise Σ is PD. Hence:

$$2\ln\left(\frac{e^b - 1}{b}\right) = 2\varphi(b) \leq \varphi(a) + \varphi(d) = \ln\left(\frac{e^a - 1}{a}\right) + \ln\left(\frac{e^d - 1}{d}\right)$$

The inequality is preserved by exponentiation:

$$\left(\frac{e^b - 1}{b}\right)^2 \leq \left(\frac{e^a - 1}{a}\right)\left(\frac{e^d - 1}{d}\right)$$

This leads at last to the inequality:

$$\left(e^b - 1\right)^2 = b^2\left(\frac{e^b - 1}{b}\right)^2 \leq \frac{b^2}{ad}\left(e^a - 1\right)\left(e^d - 1\right) = \left(\frac{b}{\gamma}\right)^2\left(e^a - 1\right)\left(e^d - 1\right) \leq 1 \cdot \left(e^a - 1\right)\left(e^d - 1\right)$$

Therefore, in this case $\left(e^b - 1\right)^2 \leq \left(e^a - 1\right)\left(e^d - 1\right)$ with equality if and only if $a = b = d$. This means

that if $b > 0$, the lognormal transformation $\Sigma \to T = e^\Sigma - 1_{2\times 2}$ is from NND to NND. But T is

NND only if $\left(e^b - 1\right)^2 = \left(e^a - 1\right)\left(e^d - 1\right)$, or only if $a = b = d$. Otherwise, T is PD. So, when

$b > 0$, $T = e^\Sigma - 1_{2\times 2}$ is PD except when all four elements of Σ have the same positive value. Even

20

the NND matrix $\Sigma = \begin{bmatrix} a & \sqrt{ad} \\ \sqrt{ad} & d \end{bmatrix}$ log-transforms into a PD matrix, unless $a = d$. So all PD and

most NND normal variances transform into PD lognormal variances. A NND lognormal variance

indicates that at least one element of the normal random vector is duplicated.

In the third and final case, $b < 0$, or more fully, $-\gamma \le b < 0$. This is equivalent to $0 < -b \le \gamma$, or

to the second case with $-b$. In that case, $(e^{-b} - 1)^2 \le (e^a - 1)(e^d - 1)$ with equality if and only if

$a = -b = d$. But from this, as well as from the fact that $0 < e^{2b} < e^{2 \cdot 0} < 1$, it follows:

$$(e^b - 1)^2 = e^{2b}(1 - e^{-b})^2 = e^{2b}(e^{-b} - 1)^2 \le e^{2b}(e^a - 1)(e^d - 1) < 1 \cdot (e^a - 1)(e^d - 1)$$

So in this case the inequality is strict: $(e^b - 1)^2 < (e^a - 1)(e^d - 1)$, and T is PD. Therefore, if $b < 0$,

the lognormal transform $\Sigma \to T = e^\Sigma - 1_{2 \times 2}$ is PD, even if Σ is NND.

To summarize, the lognormal transformation $\Sigma \to T = e^\Sigma - 1_{2 \times 2}$ is PD if Σ is PD. Even when Σ is

not PD, but merely NND, T is almost always PD. Only when Σ is so NND as to conceal a

random-variable duplication is its lognormal transformation NND.

The Hadamard (elementwise) product and Schur's product theorem allow for an understanding of

the general lognormal transformation $\Sigma \to T = e^\Sigma - 1_{n \times n}$. Denoting the elementwise n^{th} power of Σ

as $\Sigma^{\circ n} = \overbrace{\Sigma \circ \ldots \circ \Sigma}^{n \text{ factors}}$, we can express elementwise exponentiation as $e^\Sigma = \sum_{j=0}^{\infty} \Sigma^{\circ j} / j!$. So

$T = e^\Sigma - 1_{n \times n} = \sum_{j=1}^{\infty} \Sigma^{\circ j} / j!$. According to Schur's theorem (§3 of Million [2007]), the Hadamard

product of two NND matrices is NND.[27] Since Σ is NND, its powers $\Sigma^{\circ j}$ are NND, as well as the terms $\Sigma^{\circ j}/j!$. Being the sum of a countable number of NND matrices, T also must be NND.[28] But if just one of the terms of the sum is PD, the sum itself must be PD. Therefore, if Σ is PD, then T also is PD.

Now the kernel of $m \times n$ matrix A is the set of all $\mathbf{x} \in \mathfrak{R}^n$ such that $A\mathbf{x} = 0$, or $ker(A) = \{\mathbf{x} : A\mathbf{x} = 0\}$. The kernel is a linear subspace of $\mathfrak{R}^n$ and its dimensionality is $n - rank(A)$. By the Cholesky decomposition the NND matrix U can be factored as $U_{n \times n} = W'W_{n \times n}$. The quadratic form in U is $\mathbf{x}'U\mathbf{x} = \mathbf{x}'W'W\mathbf{x} = (W\mathbf{x})'(W\mathbf{x})$. If $\mathbf{x}'U\mathbf{x} = 0$, then $W\mathbf{x} = 0_{n \times 1}$, and $U\mathbf{x} = W'W\mathbf{x} = W'0_{n \times 1} = 0_{n \times 1}$. Conversely, if $U\mathbf{x} = 0_{n \times 1}$, then $\mathbf{x}'U\mathbf{x} = 0$. So the kernel of NND matrix U is precisely the solution set of $\mathbf{x}'U\mathbf{x} = 0$, i.e., $\mathbf{x}'U\mathbf{x} = 0$ if and only if $x \in ker(U)$. Therefore, the kernel of $T = \sum_{j=1}^{\infty} \Sigma^{\circ j}/j!$ is the intersection of the kernels of $\Sigma^{\circ j}$, or $ker(T) = \bigcap_{j=1}^{\infty} ker(\Sigma^{\circ j})$. It is possible for this intersection to be of dimension zero, i.e., for it to equal $\{0_{n \times 1}\}$, even though the kernel of no $\Sigma^{\circ j}$ is. Because of the accumulation of intersections in $\sum_{j=1}^{k \to \infty} \Sigma^{\circ j}/j!$, the lognormal transformation of NND matrix Σ tends to be "more PD" than Σ itself.

---

[27] Appendix C provides a proof of this theorem.

[28] For the quadratic form of a sum equals the sum of the quadratic forms: $\mathbf{x}'\left(\sum_k U_k\right)\mathbf{x} = \sum_k \mathbf{x}'U_k\mathbf{x}$. In fact, if the coefficients a_j are non-negative, then $\sum_{j=1}^{\infty} a_j \Sigma^{\circ j}$ is NND, provided that the sum converges, as e^{Σ} does.

We showed above that if Σ is PD, then $T = e^{\Sigma} - 1_{n \times n}$ is PD. But even if Σ is NND, T will be PD,

unless Σ contains a 2×2 subvariance $\begin{bmatrix} \Sigma_{jj} & \Sigma_{jk} \\ \Sigma_{kj} & \Sigma_{kk} \end{bmatrix}$ whose four elements are all equal.

APPENDIX B

THE NORMAL MOMENT GENERATING FUNCTION AND COMPLEX ARGUMENTS

In Appendix A we saw that the moment generating function of the real-valued normal multivariate

$\mathbf{x} \sim N(\mu, \Sigma)$, viz., $M_{\mathbf{x}}(t) = e^{t'\mu + t'\Sigma t/2}$, is valid at least for $t \in \Re^n$. The validity rests on the identity

$$\int_{\mathbf{x} \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-[\mu+\Sigma t])'\Sigma^{-1}(\mathbf{x}-[\mu+\Sigma t])} dV = 1 \text{ for real-valued } \xi = \mu + \Sigma t. \text{ But in Section 8 we must}$$

know the value of $\varphi(\xi) = \int_{\mathbf{x} \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\xi)'\Sigma^{-1}(\mathbf{x}-\xi)} dV$ when ξ is a complex $n \times 1$ vector. So in

this appendix, we will prove that for all $\xi \in C^n$, $\varphi(\xi) = 1$.

The proof begins with diagonalization. Since Σ is symmetric and PD, Σ^{-1} exists and is symmetric and PD. According to the Cholesky decomposition (Healy [1986, §7.2]), there exists a real-valued $n \times n$ matrix W such that $\mathbf{W}'\mathbf{W} = \Sigma^{-1}$. Due to theorems on matrix rank, W must be non-singular, or invertible. So the transformation $\mathbf{y} = \mathbf{W}\mathbf{x}$ is one-to-one. And letting $\zeta = \mathbf{W}\xi$, we have $\mathbf{y} - \zeta = \mathbf{W}(\mathbf{x} - \xi)$. Moreover, the volume element in the y coordinates is:

$$dV_{\mathbf{y}} = \|\mathbf{W}\| dV_{\mathbf{x}} = \sqrt{|\mathbf{W}|^2} dV_{\mathbf{x}} = \sqrt{|\mathbf{W}'| |\mathbf{W}|} dV_{\mathbf{x}} = \sqrt{|\mathbf{W}'\mathbf{W}|} dV_{\mathbf{x}} = \sqrt{|\Sigma^{-1}|} dV_{\mathbf{x}} = \frac{dV_{\mathbf{x}}}{\sqrt{|\Sigma|}}.$$

Hence:

$$\varphi(\xi) = \int\limits_{x \in \mathfrak{R}^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\xi)' \Sigma^{-1}(x-\xi)} dV$$

$$= \int\limits_{x \in \mathfrak{R}^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\xi)' w'w(x-\xi)} dV_x$$

$$= \int\limits_{x \in \mathfrak{R}^n} \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}[w(x-\xi)]'[w(x-\xi)]} \frac{dV_x}{\sqrt{|\Sigma|}}$$

$$= \int\limits_{y \in \mathfrak{R}^n} \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}(y-\zeta)'(y-\zeta)} dV_y$$

$$= \int\limits_{y \in \mathfrak{R}^n} \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}\sum_{j=1}^{n}(y_j-\zeta_j)^2} dV$$

$$= \int\limits_{y \in \mathfrak{R}^n} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sum_{j=1}^{n}(y_j-\zeta_j)^2} dV$$

$$= \prod_{j=1}^{n} \int\limits_{y_j=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sum_{j=1}^{n}(y_j-\zeta_j)^2} dy_j$$

$$= \prod_{j=1}^{n} \psi(\zeta_j)$$

In the last line $\psi(\zeta) = \lim\limits_{\substack{a \to -\infty \\ b \to +\infty}} \int\limits_{x=a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\zeta)^2} dx$. Obviously, if ζ is real-valued, $\psi(\zeta) = 1$. So the

issue of the value of a moment generating function of a complex variable resolves into the issue of

the "total probability" of a unit-variance normal random variable with a complex mean.[29]

To evaluate $\psi(\zeta)$ requires some complex analysis with contour integrals. First, consider the

standard-normal density function with a complex argument: $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. By function-

composition rules, since both z^2 and the exponential function are "entire" functions (i.e., analytic

---

[29] We deliberately put 'total probability' in quotes because the probability density function with complex ζ is not proper; it may produce negative and even complex values for probability densities.

over the whole complex plane), so too is $f(z)$. Therefore, $\oint_C f(z)dz = 0$ for any closed contour C

(cf. Appendix D.7 of Havil [2003]). Let C be the parallelogram traced from vertex $z = b$ to vertex

$z = a$ to vertex $z = a - \zeta$ to vertex $z = b - \zeta$ and finally back to vertex $z = b$. Therefore:

$$0 = \oint_C f(z)dz$$

$$= \int_b^a f(z)dz + \int_a^{a-\zeta} f(z)dz + \int_{a-\zeta}^{b-\zeta} f(z)dz + \int_{b-\zeta}^b f(z)dz$$

The line segments along which the second and fourth integrals proceed are finite; their common

length is $L = |(a - \zeta) - a| = |-\zeta| = |\zeta| = |b - (b - \zeta)|$, where $|\zeta| = \sqrt{\zeta \overline{\zeta}} \geq 0$. By the triangle inequality

$\left| \int_a^{a-\zeta} f(z)dz \right| \leq \int_a^{a-\zeta} |f(z)dz| = \int_a^{a-\zeta} |f(z)||dz|$. But $|f(z)|$ is a continuous real-valued function, so over a

closed interval it must be upper-bounded by some positive real number M. Hence,

$\left| \int_a^{a-\zeta} f(z)dz \right| \leq \int_a^{a-\zeta} |f(z)||dz| \leq \int_a^{a-\zeta} Sup(|f(z)|)|dz| = Sup(|f(z \in [a, a - \zeta])|) \int_a^{a-\zeta} |dz| = M(a) \cdot L$.   Likewise,

$\left| \int_{b-\zeta}^b f(z)dz \right| \leq Sup(|f(z \in [b, b - \zeta])|) \int_{b-\zeta}^b |dz| = M(b) \cdot L$.

Now, in general:

$$|f(z)| = \sqrt{f(z)\overline{f(z)}}$$

$$= \sqrt{\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \cdot \overline{\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}}}$$

$$= \frac{\sqrt{e^{-\frac{z^2}{2}} \cdot \overline{e^{-\frac{z^2}{2}}}}}{\sqrt{2\pi}}$$

$$\propto \sqrt{e^{-\frac{z^2}{2}} \cdot \overline{e^{-\frac{z^2}{2}}}}$$

$$\propto \sqrt{e^{-\frac{z^2+\overline{z^2}}{2}}}$$

$$\propto \sqrt{e^{-\mathrm{Re}(z^2)}}$$

Therefore, since $a \in \Re$:

$$\lim_{a \to -\infty} M(a) \cdot L \propto \lim_{a \to -\infty} Sup\left(\sqrt{e^{-\mathrm{Re}(z^2)}}; z \in [a, a-\zeta]\right) \cdot L$$

$$\propto \lim_{a^2 \to +\infty} Sup\left(\sqrt{e^{-a^2 \mathrm{Re}\left(\left(\frac{z}{a}\right)^2\right)}}; \frac{z}{a} \in \left[1, 1-\frac{\zeta}{a}\right]\right) \cdot L$$

$$\propto \sqrt{e^{-\infty \mathrm{Re}(1)}} \cdot L$$

$$\propto 0$$

Similarly, $\lim_{b \to +\infty} M(b) \cdot L \propto 0$. So in the limit as $a \to -\infty$ and $b \to +\infty$ on the real axis, the second

and fourth integrals approach zero. Accordingly:

$$0 = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \{0\}$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \left\{\int_b^a f(z)dz + \int_a^{a-\zeta} f(z)dz + \int_{a-\zeta}^{b-\zeta} f(z)dz + \int_{b-\zeta}^b f(z)dz\right\}$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \left\{\int_b^a f(z)dz + \int_{a-\zeta}^{b-\zeta} f(z)dz\right\}$$

And so, $\displaystyle \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{a-\zeta}^{b-\zeta} f(z)dz = -\lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_b^a f(z)dz = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f(z)dz = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$

So, at length:

$$\psi(\zeta) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{x=a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\zeta)^2} dx$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{x=a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}((z+\zeta)-\zeta)^2} d(z+\zeta)$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{z=a-\zeta}^{b-\zeta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} d(z+\zeta)$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{z=a-\zeta}^{b-\zeta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

$$= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_{z=a-\zeta}^{b-\zeta} f(z) dz$$

$$= 1$$

So, working backwards, what we proved for one dimension, viz., $\psi(\zeta) = \displaystyle\int_{x=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\zeta)^2} dx = 1$,

applies n-dimensionally: for all $\xi \in C^n$, $\varphi(\xi) = \displaystyle\int_{x \in \Re^n} \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\xi)'\Sigma^{-1}(x-\xi)} dV = 1$. Therefore, even

for complex t, $M_{\mathbf{x}}(t) = e^{t'\mu + t'\Sigma t/2}$. Complex values are allowable as arguments in the moment generating function of a real-valued normal vector. This result is critical to Section 8.

Though we believe the contour-integral proof above to be worthwhile for its instructional value, a simple proof comes from the powerful theorem of analytic continuation (cf. Appendix D.12 of Havel [2003]). This theorem concerns two functions that are analytic within a common domain. If the functions are equal over any smooth curve within the domain, no matter how short,[30] then they

---

[30] The length of the curve must be positive; equality at single points, or punctuated equality, does not qualify.

are equal over all the domain. Now $\psi(\zeta) = \lim\limits_{\substack{a \to -\infty \\ b \to +\infty}} \int\limits_{x=a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\zeta)^2} dx$ is analytic over all the complex

plane. And for all real-valued ζ, $\psi(\zeta) = 1$. So $\psi(\zeta)$ and $f(\zeta) = 1$ are two functions analytic over the

complex plane and identical on the real axis. Therefore, by analytic continuation $\psi(\zeta)$ must equal

one for all complex ζ. Analytic continuation is analogous with the theorem in real analysis that two

smooth functions equal over any interval are equal everywhere. Analytic continuation derives from

the fact that a complex derivative is the same in all directions. It is mistaken to regard the real and

imaginary parts of the derivative as partial derivatives, as if they applied respectively to the real and

imaginary axes of the independent variable. Rather, the whole derivative applies in every direction.

APPENDIX C

EIGEN-DECOMPOSITION AND SCHUR'S PRODUCT THEOREM

Appendix A quoted Schur's Product Theorem, viz., that the Hadamard product of non-negative-definite (NND) matrices is NND. Million [2007] proves it as Theorem 3.4; however, we believe our proof in this appendix to be simpler; moreover, it affords a review of eigen-decomposition. Those familiar with eigen-decomposition may skip to the last paragraph.

Let Γ be an $n \times n$ Hermetian NND matrix. As explained in Section 4, 'Hermetian' means that $\Gamma = \Gamma^*$; 'NND' means that for every complex $n \times n$ vector z (or $z \in C^n$), $z^* \Gamma z \geq 0$. Positive definiteness [PD] is a stricter condition, in which $z^* \Gamma z = 0$ if and only if $z = 0_{n \times 1}$.

Complex scalar λ and non-zero vector v form an "eigenvalue-eigenvector" pair of Γ, if $\Gamma v = \lambda v$. Since $v = 0_{n \times 1}$ is excluded as a trivial solution, vector v can be scaled to unity, or $v^* v = 1$. But $\Gamma v = \lambda v$ if and only if $(\Gamma - \lambda I_n)v = 0_{n \times 1}$. If $\Gamma - \lambda I_n$ is non-singular, or invertible, then:

$$v = I_n v = (\Gamma - \lambda I_n)^{-1}(\Gamma - \lambda I_n)v = (\Gamma - \lambda I_n)^{-1} 0_{n \times 1} = 0_{n \times 1}$$

Hence, allowable eigenvectors require for $\Gamma - \lambda I_n$ to be singular, or for its determinant $|\Gamma - \lambda I_n|$ to be zero. Since the determinant is an nth-degree equation (with complex coefficients based on the elements of Γ) in λ, it has n root values of λ, not necessarily distinct. So the determinant can be factored as $f(\lambda) = |\Gamma - \lambda I_n| = (\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$. Since $f(\lambda_j) = |\Gamma - \lambda_j I_n| = 0$, there exist non-zero solutions to $(\Gamma - \lambda I_n)v = 0_{n \times 1}$. So for every eigenvalue there is a non-zero eigenvector, even a non-zero eigenvector of unit magnitude.

The first important result is that the eigenvalues of Γ are real-valued and non-negative. Consider the jth eigenvalue-eigenvector pair, which satisfies the equation $\Gamma v_j = \lambda_j v_j$. Therefore, $v_j^* \Gamma v_j = \lambda_j v_j^* v_j$. Since Γ is NND, $v_j^* \Gamma v_j$ is real-valued and non-negative. Also, $v_j^* v_j$ is real-valued and positive. Therefore, their quotient λ_j is a real-valued and non-negative scalar. Furthermore, if Γ is PD, $v_j^* \Gamma v_j$ is positive, as well as λ_j.

The second important result is that eigenvectors paired with unequal eigenvalues are orthogonal. Let the two unequal eigenvalues be $\lambda_j \neq \lambda_k$. Because the eigenvalues are real-valued, $\overline{\lambda}_j = \lambda_j$. The eigenvector equations are $\Gamma v_j = \lambda_j v_j$ and $\Gamma v_k = \lambda_k v_k$. The following string of equations relies on Γ's being Hermetian (so $\Gamma = \Gamma^*$):

$$
\begin{aligned}
\left(\lambda_j - \lambda_k\right) v_j^* v_k &= \lambda_j v_j^* v_k - \lambda_k v_j^* v_k \\
&= \overline{\lambda}_j v_j^* v_k - \lambda_k v_j^* v_k \\
&= \left(\lambda_j v_k^* v_j\right)^* - \lambda_k v_j^* v_k \\
&= \left(v_k^* \Gamma v_j\right)^* - v_j^* \Gamma v_k \\
&= v_j^* \Gamma^* v_k - v_j^* \Gamma v_k \\
&= v_j^* \Gamma v_k - v_j^* \Gamma v_k \\
&= 0
\end{aligned}
$$

Because $\lambda_j - \lambda_k \neq 0$, the eigenvectors must be orthogonal, or $v_j^* v_k = 0$. If all the eigenvectors are distinct, the eigenvectors form an orthogonal basis of C^n. But even if not, the kernel of each eigenvalue, or $ker\left(\Gamma - \lambda_j I_n\right) = \left\{ z \in C^n : \Gamma z = \lambda_j z \right\}$ is a linear subspace of C^n whose rank or dimensionality equals the multiplicity of the root λ_j in the characteristic equation $f(\lambda) = \left|\Gamma - \lambda I_n\right|$. This means that the number of mutually orthogonal eigenvectors paired with an eigenvalue equals

how many times that eigenvalue is a root of its characteristic equation. Consequently, there exist n eigenvalue-eigenvector pairs $(\lambda_j, \mathbf{v}_j)$ such that $\Gamma\mathbf{v}_j = \lambda_j\mathbf{v}_j$ and $\mathbf{v}_j^*\mathbf{v}_k = \delta_{ij}$.[31]

Now, define W as the partitioned matrix $\mathbf{W}_{n\times n} = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]$. The jk^{th} element of $\mathbf{W}^*\mathbf{W}$ equals $\mathbf{v}_j^*\mathbf{v}_k = \delta_{ij}$; hence, $\mathbf{W}^*\mathbf{W} = \mathbf{I}_n$. A matrix whose transjugate is its inverse is called "unitary," as is W.[32] Furthermore, define Λ as the $n\times n$ diagonal matrix whose jj^{th} element is λ_j. Then:

$$\Gamma\mathbf{W} = \Gamma[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n] = [\Gamma\mathbf{v}_1 \quad \cdots \quad \Gamma\mathbf{v}_n] = [\lambda_1\mathbf{v}_1 \quad \cdots \quad \lambda_n\mathbf{v}_n] = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \mathbf{W}\Lambda$$

And so, $\Gamma = \Gamma\mathbf{I}_n = \Gamma\mathbf{W}\mathbf{W}^* = \mathbf{W}\Lambda\mathbf{W}^*$, and Γ is said to be "diagonalized." Thus have we shown, assuming the theory of equations, [33] the third important result, viz., that every NND Hermetian matrix can be diagonalized. Other matrices can be diagonalized; the NND [or PD] consists in the fact that all the eigenvalues of this diagonalization are non-negative [or positive].

The fourth and final "eigen" result relies on the identity $\mathbf{W}^*\mathbf{v}_j = \mathbf{e}_j$, which just extracts the jth columns of each side of $\mathbf{W}^*\mathbf{W} = \mathbf{I}_n$. As in Appendix A, $\mathbf{e}_j$ is the jth unit vector. Therefore:

---

[31] The Kronecker delta, δ_{ij}, is the function $\text{IF}(i = j, 1, 0)$.

[32] To be precise, at this point W* is only the left-inverse of W. But by matrix-rank theorems, the rank of W equals n, so W has a unique full inverse $\mathbf{W}^{-1}$. Then $\mathbf{W}^* = \mathbf{W}^*\mathbf{I}_n = \mathbf{W}^*(\mathbf{W}\mathbf{W}^{-1}) = (\mathbf{W}^*\mathbf{W})\mathbf{W}^{-1} = \mathbf{I}_n\mathbf{W}^{-1} = \mathbf{W}^{-1}$.

[33] The theory of equations guarantees the existence of the roots of the nth-degree equation $f(\lambda) = |\Gamma - \lambda\mathbf{I}_n|$. Appendix D.9 of Havel [2003] contains a quick and lucid proof of this, the fundamental theorem of algebra.

$$\Gamma = \mathrm{W}\Lambda\mathrm{W}^*$$

$$= \mathrm{W}\left(\sum_{j=1}^{n}\lambda_j\mathrm{e}_j\mathrm{e}_j^*\right)\mathrm{W}^*$$

$$= \mathrm{W}\left(\sum_{j=1}^{n}\lambda_j\left(\mathrm{W}^*\mathrm{v}_j\right)\left(\mathrm{W}^*\mathrm{v}_j\right)^*\right)\mathrm{W}^*$$

$$= \mathrm{W}\left(\sum_{j=1}^{n}\lambda_j\mathrm{W}^*\mathrm{v}_j\mathrm{v}_j^*\mathrm{W}\right)\mathrm{W}^*$$

$$= \mathrm{W}\mathrm{W}^*\left(\sum_{j=1}^{n}\lambda_j\mathrm{v}_j\mathrm{v}_j^*\right)\mathrm{W}\mathrm{W}^*$$

$$= \mathrm{I}_n\left(\sum_{j=1}^{n}\lambda_j\mathrm{v}_j\mathrm{v}_j^*\right)\mathrm{I}_n$$

$$= \sum_{j=1}^{n}\lambda_j\mathrm{v}_j\mathrm{v}_j^*$$

The form $\sum_{j=1}^{n}\lambda_j\mathrm{v}_j\mathrm{v}_j^*$ is called the "spectral decomposition" of Γ (§7.4 of Healy [1986]), which plays

the leading role in the following succinct proof of Schur's product theorem.

If Σ and T are two $n{\times}n$ Hermetian NND definite matrices, we may spectrally decompose them as

$\Sigma = \sum_{j=1}^{n}\lambda_j\mathrm{v}_j\mathrm{v}_j^*$ and $T = \sum_{j=1}^{n}\kappa_j\eta_j\eta_j^*$, where all the λ and $\varkappa$ scalars are non-negative. Accordingly:

$$(\Sigma \circ T)_{jk} = (\Sigma)_{jk} (T)_{jk}$$

$$= \left(\sum_{r=1}^{n} \lambda_r v_r v_r^* \right)_{jk} \left(\sum_{s=1}^{n} \kappa_s \eta_s \eta_s^* \right)_{jk}$$

$$= \left(\sum_{r=1}^{n} \lambda_r (v_r)_j (\overline{v}_r)_k \right) \left(\sum_{s=1}^{n} \kappa_s (\eta_s)_j (\overline{\eta}_s)_k \right)$$

$$= \sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r \kappa_s (v_r)_j (\eta_s)_j (\overline{v}_r)_k (\overline{\eta}_s)_k$$

$$= \sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r \kappa_s (v_r \circ \eta_s)_j (\overline{v}_r \circ \overline{\eta}_s)_k$$

$$= \sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r \kappa_s (v_r \circ \eta_s)_j \overline{(v_r \circ \eta_s)}_k$$

$$= \sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r \kappa_s \left((v_r \circ \eta_s)(v_r \circ \eta_s)^* \right)_{jk}$$

$$= \left(\sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r \kappa_s (v_r \circ \eta_s)(v_r \circ \eta_s)^* \right)_{jk}$$

Hence, $\Sigma \circ T = \sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r \kappa_s (v_r \circ \eta_s)(v_r \circ \eta_s)^*$. Since each matrix $(v_r \circ \eta_s)(v_r \circ \eta_s)^*$ is Hermetian

NND, and each scalar $\lambda_r \kappa_s$ is non-negative, $\Sigma \circ T$ must be Hermetian NND. Therefore, the

Hadamard product of NND matrices is NND.