Exercise*

(1) Prove that for a random vector X with the standard $N(0,\Sigma)$ pdf, Σ is in fact the covariance matrix of X. [Suggestion: First prove it in the univariate case, then in the case of independent X, then in the general case, by writing a general X as $W'Z$ where Z is independent.]

For the univariate case, we can apply integration by parts:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}}\,dx = \frac{1}{\sigma\sqrt{2\pi}}\left(-\sigma^2 x e^{-\frac{x^2}{2\sigma^2}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\sigma^2 e^{-\frac{x^2}{2\sigma^2}}\,dx\right) = 0 + \sigma^2\cdot 1$$

For a vector X of n independent $N(0,\sigma_i^2)$ variables, with $i = 1,\dots,n$, the pdf is a product of univariate normal pdfs. When we integrate x_i^2 over this joint pdf, it is easy to see that the univariate argument above tells us that σ_i^2 is the variance. The off-diagonal elements of Σ here are 0, and the covariance of X_i with X_j for $i \neq j$ is obtained by integrating $x_i x_j$ with respect to the joint pdf. It is easy to see that this breaks into a product of separate integrals that are 1 when we integrate with respect to $k \neq j, i$ and 0 when we integrate with respect to i or j, because x_i is anti-symmetric about zero as a function of x_i, while the normal pdf itself is itself symmetric about zero. We know from our discussion of matrix transformations that any $N(0,\Sigma)$ variable can be written as $W'Z$, where $W'W = \Sigma$ and Z is multivariate normal with covariance matrix I. But we also know that if $\mathrm{Var}(Y) = \Omega$ and C is a constant matrix, $\mathrm{Var}(CY) = C\Omega C'$. So we can conclude that $\mathrm{Var}(X) = \mathrm{Var}(W'Z) = W'W = \Sigma$.

(2) Construct an example of a three-dimensional random vector X with the property that each of the three possible two-element subvectors (X_i, X_j) of X has a marginal distribution that is $N(0, I)$, but the three X's considered jointly are not independent. [Though this is in some sense simple, it may be hard to see how to approach it. Don't waste a lot of time on it if it seems impossible.] *Here's one way to do this: Take $(X,Y,Z) \sim N(0,I)$, and then form the conditional distribution of these three variables given $XYZ > 0$. This conditional distribution has joint pdf $2\phi(x)\phi(y)\phi(z)$ on those regions in $\mathbb{R}^3$ where $xyz > 0$. This is because the pdf is proportional to the original pdf in those regions, and those regions,*

$$\{x,y,z \mid x > 0, y > 0, z > 0\} \cup \{x,y,z \mid x > 0, y < 0, z < 0\} \\ \cup \{x,y,z \mid x < 0, y > 0, z < 0\} \cup \{x,y,z \mid x < 0, y < 0, z > 0\}\ ,$$

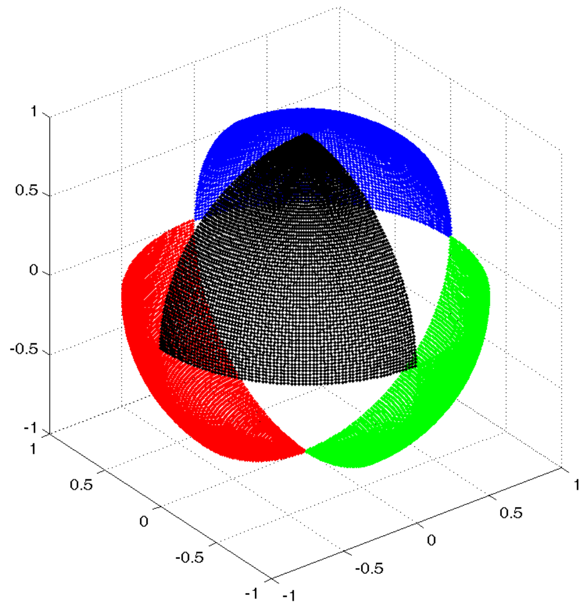

FIGURE 1. The four components of a level surface of the joint pdf of X, Y, Z

are four of the eight similar orthants that make up $\mathbb{R}^3$. The pdf is zero in the remainder of the space. But then if we integrate the pdf with respect to z for a fixed x, y, say with $xy > 0$, we will be forming

$$\int_0^\infty 2\phi(x)\phi(y)\phi(z)\,dz = \phi(x)\phi(y)\,.$$

There is a similar expression for the case where $xy < 0$, though in that case we integrate over the region where $z < 0$. Still, in either case, the marginal pdf emerges as simply $\phi(x)\phi(y)$, which is the 2-dimensional $N(0, I)$ pdf. Thus the pair X, Y is independent and jointly normal considered by itself, but does not form a joint normal vector with Z. That the 3-d joint distribution is not independent follows from the fact that the marginal pdf's are all everywhere positive, so their product can never be zero, yet the joint pdf is in fact zero over half the space. Figure 1 shows the four nontrivial pieces of one level surface of the pdf. Note that it is just four chunks of a sphere.

(3) Using a Taylor approximation of the log of the pdf about its maximum, construct normal approximations to the $\Gamma(1.5)$, $\Gamma(2)$, and $\Gamma(4)$ pdf's. In each case, plot both the original pdf and the normal approximation to it. Do this also for $\Gamma(.5)$, though in this case, since the pdf does not have a peak, you should

make the Taylor expansion about the mean (which is .5) and, because this is not the peak, you will have a first-order as well as a second-order term.

(No answer to this one, since apparently nearly everyone could see how to do it.)

(4) Here are two Σ matrices:

$$\begin{bmatrix} 1.1000 & 0.1000 & 0.1000 & 1.0000 \\ 0.1000 & 1.1000 & 1.0000 & 0.1000 \\ 0.1000 & 1.0000 & 1.1000 & 0.1000 \\ 1.0000 & 0.1000 & 0.1000 & 1.1000 \end{bmatrix} \begin{bmatrix} 1.0000 & 0.5000 & 0.3333 & 0.2500 \\ 0.5000 & 0.2750 & 0.1667 & 0.1250 \\ 0.3333 & 0.1667 & 0.1222 & 0.0833 \\ 0.2500 & 0.1250 & 0.0833 & 0.0688 \end{bmatrix}.$$

For each, compute both an eigenvalue decomposition and a Choleski decomposition. Do both methods "work" to suggest structure for the matrix? Do they suggest similar definitions of the important "Z" components explaining variation in the X's with these covariance matrices? Matlab commands relevant here are `chol` and `eig`.

For the first matrix, the Choleski factor is

$$W = \begin{bmatrix} 1.0488 & 0.0953 & 0.0953 & 0.9535 \\ 0 & 1.0445 & 0.9487 & 0.0087 \\ 0 & 0 & 0.4368 & 0.0019 \\ 0 & 0 & 0 & 0.4368 \end{bmatrix}$$

With $X = W'Z$, we can see that the first two elements of the Z vector have large weight (large entries in the first two rows of W) for every X_i. The variance of X_i is the sum of squared elements in the i'th row, and the variance of the part made up of Z_1 and Z_2 alone is the sum of squares of the first two elements in the i'th row. So a convenient summary of how much of each X_i we explain with the first two Z's is found by

```
W2=(W.*W)';
sum(W2(1:2,:))./sum(W2);
```

which produces the vector [1 1 .8265 .8265]. So the first two X's are just linear combinations of the first two Z's, while the last two have other variation as well. They are mostly accounted for by the first two Z's, but maybe the principal components W will do better.

The principal components W is (I'm skipping details, since most people figured out how to get Matlab to do these calculations)

$$W = \begin{bmatrix} -0.0177 & 0.2229 & -0.2229 & 0.0177 \\ -0.2229 & -0.0177 & 0.0177 & 0.2229 \\ 0.6892 & -0.6892 & -0.6892 & 0.6892 \\ 0.7583 & 0.7583 & 0.7583 & 0.7583 \end{bmatrix}.$$

This too has only two big rows, but in this case they are the last two. The same sort of calculation as above, but this time summing over the last two rows of W2 in the numerator instead of the first two, produces the vector [.9545 .9545 .9545 .9545]. This seems to do a neater job of extracting common components here than the Choleski factor.

Repeating these calculations with the other matrix, in which a single row turns out to dominate, one obtains the vectors [1 .9091 .9091 .9084] (Choleski) and [.9960 .9381 .9198 .9128] (principal components). Once again, principal components seems to have found a single "common factor" that explains more of the other variables' variation, while sacrificing only slightly on explanation of the "first" variable. An interesting question is whether the two methods are finding nearly the same "big Z" in this latter matrix. Since these are two normal random variables, there is a regression relation connecting them, $Z_{chol} = \beta Z_{pc} + \varepsilon$, and we know how to calculate β and the variance of ε, because we can calculate the variances and covariance of the two Z vectors. We can calculate them because both Z vectors are linear functions of X, via the relation $\left(W^{-1}\right)' X = Z$. So if the Choleski "big row" in W is the first, we take $c = W^{\cdot 1}$, the first column of W^{-1}. When we similarly pull out the column of the principal components W^{-1} that corresponds to the big column of W and label it b. The variances of the two Z's are then $c'\Sigma c$ and $b'\Sigma b$ (both of which are by construction 1), and the covariance is $c'\Sigma b$. We get in this case $b =$[.6983 .3554 .2346 .1753] and $c =$[1 0 0 0]. This gives us the regression equation

$$Z_{chol} = .9980 Z_{pc} + \varepsilon\,.$$

Since both Z's have variance 1, it is clear that the two common components are nearly the same. In fact the proportion of the variance of the Choleski Z that is accounted for by the principal components Z is $.9980^2 = .9960$. This measure of "explained" variance as a fraction of total variance of the left-hand side variable is what is known as the R^2 of the regression.