Dave Backus @ NYU

Notes on Factor Models*

(Started: October 31, 2011; Revised: February 18, 2014)

This pulls together some thoughts on topics related to factor models. The idea throughout is to reduce the dimension of the data and consider the various issues that come up when dealing with latent factors.

http://fxdiebold.blogspot.com/2014/02/thoughts-on-factor-augmented-vars.html

1 Transformations and rotations of multivariate normals

There are some common issues referred to as identification, but they're typically closer to normalizations. Here's the simplest example I could think of: an n-dimensional random vector $y \sim \mathcal{N}(0, V)$ whose covariance matrix $V = E(yy^\top)$ has rank n. The question is whether we can connect y to a vector w of independent standard normals.

In fact, we can do this in more than one way. The Cholesky decomposition gives us $V = LL^\top$ with L lower triangular. Then $y = Lw$ has covariance V and $w = L^{-1}y$. We can also use any orthogonal transformation of w. For example, let $y = LQw$, with Q an orthogonal matrix ($Q^\top Q = QQ^\top = I$) Then $E(yy^\top) = (LQ)(LQ)^\top = LL^\top = V$. The transformation tends to kill the lower triangularity, but it shows ut that we have lots of options here.

Another example is the spectral decomposition: $V = QDQ^\top$ with Q (again) orthogonal and D is diagonal. Here Q's columns are the eigenvectors of V and the elements of D are the eigenvalues. If we define $T = QD^{1/2}$, then $y = Tw$ has covariance V. We can find the w's indirectly from $w = T^{-1}y$.

Both decompositions reduce the number of parameters. If we observe y, then we can estimate V. Since V is symmetric, that gives us $n(n+1)/2$ pieces of information. So we can't estimate more parameters than that. In the Cholesky decomposition, we kill off the extra parameters by zeroing out the elements above the diagonal. In the spectral decomposition, we accomplish the same thing with othogonality conditions among the columns.

2 Factor analysis

A lot of this is adapted from Anderson and Rubin, "Statistical inference in factor analysis," 1956, which is a thing of beauty.

Consider an n-dimensional vector of data y ("indicators") with mean zero and variance $E(yy^\top) = V_y$. The idea of factor analysis is to approximate y with a vector x ("factors") of dimension $k < n$:

$$y \quad = \quad Bx + u. \tag{1}$$

*Working notes, no guarantee of accuracy or sense.

Here $E(xx^\top) = V_x$, $E(uu^\top) = D$ is diagonal, and (x, u) are independent. That implies a restricted variance matrix

$$V_y \;=\; BE(xx^\top)B^\top + D \;=\; BV_xB^\top + D. \tag{2}$$

Special case: $D = 0$.

The question again is whether we can identify the x's and the parameters that connect them to the y's.

This places some structure on covariance matrix, which you can see by counting parameters. S has (since it's symmetric) $n(n+1)/2$ parameters. The rhs has $nq+n$, which is smaller if $q < (n-1)/2$. This overstates the rhs, because for large q, $BB^\top$ doesn't need nq parameters. [What's the logic here?]

3 Principal components

Take the same starting point: a vector y with mean zero and variance S. The spectral decomposition gives us

$$S \;=\; QDQ^\top \;=\; d_1v_1v_1^\top + d_2v_2v_2^\top + \cdots + d_nv_nv_n^\top,$$

where v_j is the jth column of Q. If we rank the eigenvalues d_j from largest to smallest, we can use (say) the first k to approximate S:

$$S \;\approx\; d_1v_1v_1^\top + d_2v_2v_2^\top + \cdots + d_kv_kv_k^\top.$$

We can think of the v's as factors. Where factor analysis gives us $S = \text{factors} + \text{diagonal}$, this gives us $S = \text{factors} + \text{small}$.

[need to link up notation with what follows]

We can also interpret principal components as approximations to the data matrix Y, an m by n matrix whose rows are observations of $y^\top$. The singular value decomposition gives us

$$Y \;=\; Q_1\Sigma Q_2^\top.$$

If we estimate S by $Y^\top Y$ (ignore for now division by number of observations), we have $\Sigma = Y^\top Y = Q_2\Sigma^\top\Sigma Q_2^\top$. That gives us the earlier decomposition with $D = \Sigma^\top\Sigma$ and $Q = Q_2$.

Now consider the expansion

$$Y \;=\; u_1\sigma_1v_1^\top + u_2\sigma_2v_2^\top + \cdots + u_n\sigma_nv_n^\top.$$

Let's truncate after k terms. That involves dropping columns of Σ and the associated rows of $Q_2^\top$. So the truncated versions: $\widehat{\Sigma}$ is m by k and $\widehat{Q}_2^\top$ is k by n. The rows of $\widehat{Q}_2^\top$ are the factors. The loadings are the diagonal matrix $\widehat{\Sigma}$, so we've saved some parameters. I think this gives us a basis for a clear count of identifiable parameter values, but I haven't worked it out yet. We could do the same with the version based on S.

A Appendix on linear algebra results

Orthogonal matrix. We say Q is orthogonal (or sometimes, orthonormal) if it's square and $Q^\top Q = QQ^\top = I$.

QR decomposition. If A is m by n with $m \geq m$, it can be factored into $A = QR$ with Q m by n and $Q^\top Q = I$ and R upper triangular. The space spanned by the columns of A is the same as that spanned by the columns of Q.

A more complete version:

$$A \quad = \quad QR \quad = \quad [Q_1 \; Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

with Q m by m orthogonal and R upper triangular. What we had before was the "thin" factorization $A = QR = Q_1 R_1$.

Spectral decomposition. Any real symmetric matrix A can be factored into $Q\Lambda Q^\top$. Here Λ is a diagonal matrix containing the (real) eigenvalues of A and Q is orthogonal with columns equal to the eigenvectors. See Strang, *Linear Algebra and Its Applications 3e*, Section 5.5. Using "column times row multiplication" (Strang problem 1.4.21), we have

$$A \quad = \quad Q\Lambda Q^\top \quad = \quad \lambda_1 v_1 v_1^\top + \lambda_2 v_2 v_2^\top + \cdots + \lambda_n v_n v_n^\top,$$

where v_j is the jth column of Q and n is the dimension of A. If A is positive definite, the eigenvalues are positive.

Singular value decomposition. Taken from Strang, *Linear Algebra and Its Applications 3e*, Appendix A. Let A be m by n. Then it can be factored into $A = Q_1 \Sigma Q_2^\top$ with Q_1, Q_2 orthonormal and Σ diagonal (not necessarily square) whose elements are called "singular values." Matching up dimensions, Q_1 is n by n, Q_2 is m by m, and Σ is n by m.

Examples and remarks:

- Strang's Example 1:

$$A \quad = \quad \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

- Strang's Example 2:

$$A \quad = \quad \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} \quad = \quad \begin{bmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} [1].$$

- Symmetric case. We get $A = Q\Lambda Q^\top$, the spectral decomposition.

- Matrix and eigenvalue connections:

$$AA^\top = Q_1 \Sigma \Sigma^\top Q_1^\top, \quad A^\top A = Q_2 \Sigma^\top \Sigma Q_2^\top.$$

  Eigenvalues are squares of those of A. More precisely: if λ is an eigenvalue of $A^\top A$ with eigenvalue v, then either (i) λ is an eigenvalue of $AA^\top$ with eigenvalue Av or (ii) $\lambda = 0$ and $Av = 0$.

- Dimension reduction 1. If an element of Σ is zero, we can just skip that dimension and reduce the size of the relevant matrices. An example is this result from Brillinger (exercise 3.10.36, p 87). Let A be j by k with rank l. Then we can write $A = BC$ for B j by l and C l by k. Think of C as the product of Σ and $Q_2^\top$, appropriately simplified.

- Dimension reduction 2. Suppose you want to approximate a high-dimensional A. If the colums of Q_1 are u_j and those of Q_2 are v_j (ie, the rows of $Q_2^\top$), then "column times row multiplication" gives us

$$A = Q_1 \Sigma Q_2^\top = u_1 \sigma_1 v_1^\top + u_2 \sigma_2 v_2^\top + \cdots + u_m \sigma_m v_m^\top.$$

  We can rank these by σ_j and take only those with large values, thus decreasing the amount of information we need to keep track of.

- Space spanned by (columns of) A same as that spanned by Q_1.

- Postmultiply by Q_2, get $AQ_2 = Q_1 \Sigma$, scaled columns of Q_1.