STAT 665 HW #3

Maggie Buffum

March 7, 2019

Problem 6.2

Prove the identities (6.1.4):

$$\begin{aligned}
\bar{\mathbf{x}}_{m+1} &= \frac{1}{m+1}(m\bar{\mathbf{x}}_m + \mathbf{x}_{m+1}), \qquad \text{and} \\
\mathbf{S}_{m+1} &= \mathbf{S}_m + \frac{m}{m+1}(\mathbf{x}_{m+1} - \bar{\mathbf{x}}_m)(\mathbf{x}_{m+1} - \bar{\mathbf{x}}_m)'
\end{aligned}$$

To show that $\bar{x}_m = \frac{1}{m+1}(m\bar{\mathbf{x}}_m + \mathbf{x}_{m+1})$, we first recall that $\bar{x}_m = \frac{1}{m}\sum_{i=1}^{m} x_i$. Consider $\bar{x}_{m+1}$:

$$\begin{aligned}
\bar{x}_{m+1} &= \frac{1}{m+1}\sum_{i=1}^{m+1} x_i \\
&= \frac{1}{m+1}\left(\sum_{i=1}^{m} x_i + x_{m+1}\right) \\
&= \frac{1}{m+1}\left(m\bar{x}_m + x_{m+1}\right)
\end{aligned}$$

Now let's show that $\mathbf{S}_{m+1} = \mathbf{S}_m + \frac{m}{m+1}(\mathbf{x}_{m+1} - \bar{\mathbf{x}}_m)(\mathbf{x}_{m+1} - \bar{\mathbf{x}}_m)'$:

$$\begin{aligned}
\mathbf{S}_{m+1} &= \sum_{i=1}^{m+1}(x_i - \bar{x}_{m+1})(x_i - \bar{x}_{m+1})' \\
&= \sum_{i=1}^{m}(x_i - \bar{x}_{m+1})(x_i - \bar{x}_{m+1})' + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \sum_{i=1}^{m}(x_i - \bar{x}_m + \bar{x}_m - \bar{x}_{m+1})(x_i - \bar{x}_m + \bar{x}_m - \bar{x}_{m+1})' + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \sum_{i=1}^{m}\left[(x_i - \bar{x}_m) + (\bar{x}_m - \bar{x}_{m+1})\right]\left[(x_i - \bar{x}_m) + (\bar{x}_m - \bar{x}_{m+1})\right]' + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \sum_{i=1}^{m}\left[(x_i - \bar{x}_m) + (\bar{x}_m - \bar{x}_{m+1})\right]\left[(x_i - \bar{x}_m)' + (\bar{x}_m - \bar{x}_{m+1})'\right] + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \sum_{i=1}^{m}(x_i - \bar{x}_m)(x_i - \bar{x}_m)' + \sum_{i=1}^{m}(x_i - \bar{x}_m)(\bar{x}_m - \bar{x}_{m+1})' + \sum_{i=1}^{m}(\bar{x}_m - \bar{x}_{m+1})\left[(x_i - \bar{x}_m)' + (\bar{x}_m - \bar{x}_{m+1})'\right] \\
&\quad + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \mathbf{S}_m + \left[\sum_{i=1}^{m}(x_i - \bar{x}_m)\right](\bar{x}_m - \bar{x}_{m+1})' + \sum_{i=1}^{m}(\bar{x}_m - \bar{x}_{m+1})(x_i - \bar{x}_{m+1})' + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \mathbf{S}_m + (m\bar{x}_m - m\bar{x}_m)(\bar{x}_m - \bar{x}_{m+1})' + (\bar{x}_m - \bar{x}_{m+1})\sum_{i=1}^{m}(x_i - \bar{x}_{m+1})' + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \mathbf{S}_m + (\bar{x}_m - \bar{x}_{m+1})(m\bar{x}_m - m\bar{x}_{m+1})' + (x_{m+1} - \bar{x}_{m+1})(x_{m+1} - \bar{x}_{m+1})' \\
&= \mathbf{S}_m + m(\bar{x}_m - \frac{1}{m+1}(m\bar{x}_m + x_{m+1}))(\bar{x}_m - \frac{1}{m+1}(m\bar{x}_m + x_{m+1}))' \\
&\quad + (x_{m+1} - \frac{1}{m+1}(m\bar{x}_m + x_{m+1}))(x_{m+1} - \frac{1}{m+1}(m\bar{x}_m + x_{m+1}))' \\
&= \mathbf{S}_m + \frac{m}{(m+1)^2}(\bar{x}_m - x_{m+1})(\bar{x}_m - x_{m+1})' + \frac{m^2}{(m+1)^2}(x_{m+1} - \bar{x}_m)(x_{m+1} - \bar{x}_m)' \\
&= \mathbf{S}_m + \frac{m}{(m+1)^2}(-1)^2(x_{m+1} - \bar{x}_m)(x_{m+1} - \bar{x}_m)' + \frac{m^2}{(m+1)^2}(x_{m+1} - \bar{x}_m)(x_{m+1} - \bar{x}_m)' \\
&= \mathbf{S}_m + \frac{m + m^2}{(m+1)^2}(x_{m+1} - \bar{x}_m)(x_{m+1} - \bar{x}_m)' \\
&= \mathbf{S}_m + \frac{m}{m+1}(x_{m+1} - \bar{x}_m)(x_{m+1} - \bar{x}_m)'
\end{aligned}$$

Problem 6.4

Suppose $\mathbf{S}_{N_1} \sim W_k(\Sigma, N_1 - 1)$ and is independent of $\mathbf{S}_{N_2} \sim W_k(\Sigma, N_2 - 1)$. For any $\mathbf{a} = (a_1, a_2, \ldots, a_k)'$, find the distribution of $\mathbf{a}'\mathbf{S}_{N_1}\mathbf{a}/\mathbf{a}'\mathbf{S}_{N_2}\mathbf{a}$.

From Example 6.1.1 we have that $\mathbf{a}'\mathbf{S}_N\mathbf{a}/\mathbf{a}'\Sigma\mathbf{a} \sim \chi^2_{N-1}$. Note that since the first parameter in each wishart random variable, we have that

$$\frac{\mathbf{a}'\mathbf{S}_{N_1}\mathbf{a}}{\mathbf{a}'\mathbf{S}_{N_2}\mathbf{a}} = \frac{\mathbf{a}'\mathbf{S}_{N_1}\mathbf{a}/\mathbf{a}'\Sigma\mathbf{a}}{\mathbf{a}'\mathbf{S}_{N_2}\mathbf{a}/\mathbf{a}'\Sigma\mathbf{a}}$$

is now a ratio of two chi-squared random variables. Consider multiplying and dividing each chi-squared random variable by its degrees of freedom. Then we have a scaled F-distribution:

$$= \frac{(N_1-1)}{(N_2-1)} \frac{\mathbf{a}'\mathbf{S}_{N_1}\mathbf{a}/(N_1-1)}{\mathbf{a}'\mathbf{S}_{N_2}\mathbf{a}/(N_2-1)} \sim \frac{(N_1-1)}{(N_2-1)} F_{(N_1-1,N_2-1)}$$

Problem 6.6

The following table gives the means and the variance-covariance matrix of four variables X_0, X_1, X_2, and X_3:

	X_0	X_1	X_2	X_3	Means
X_0	60.516	0.998	3.511	21.122	18.3
X_1	0.998	15.129	23.860	1.793	14.9
X_2	3.511	23.860	54.756	3.633	30.5
X_3	21.122	1.793	3.633	18.225	7.8

(a) Compute the multiple correlation coefficient of X_0 on X_1, X_2 and X_3.

Given that in part (b) we want to find the partial correlation between X_0 and X_3, I'm going to rearrange the variance-covariance matrix as

	X_0	X_3	X_2	X_1
X_0	60.516	21.122	3.511	0.998
X_3	21.122	18.225	3.633	1.793
X_2	3.511	3.633	54.756	23.860
X_1	0.998	1.793	23.860	15.129

Let's partition the matrix such that

```
S_00 <- as.matrix(cov_mat[1,1])
s_01 <- t(as.matrix(cov_mat[1,2:4]))
s_10 <- as.matrix(cov_mat[2:4,1])
S_1  <- as.matrix(cov_mat[2:4,2:4])

S_00
```

```
##        [,1]
## [1,] 60.516
```

```
s_01
```

```
##         X_3   X_2   X_1
## [1,] 21.122 3.511 0.998
```

```
s_10
```

```
##       [,1]
## X_3 21.122
## X_2  3.511
## X_1  0.998
```

```
S_1
```

```
##        X_3    X_2    X_1
## X_3 18.225  3.633  1.793
## X_2  3.633 54.756 23.860
```

```
## X_1  1.793 23.860 15.129
```

Using Result 6.2.5, we can calculate the $\hat{\rho}_{0(1,\ldots,k)}$ as

```
rho_hat <- (1/(S_00)^(1/2)) %*% (s_01 %*% solve(S_1) %*% s_10)^(1/2)
rho_hat
```

```
##           [,1]
## [1,] 0.6377814
```

(b) Compute the partial correlation coefficient between X_0 and X_3, eliminating the effects of X_1 and X_2.

Consider now that we partition the variance-covariance matrix as

$$\left[\begin{array}{cc|cc} 60.516 & 21.122 & 3.511 & 0.998 \\ 21.122 & 18.225 & 3.633 & 1.793 \\ \hline 3.511 & 3.633 & 54.756 & 23.860 \\ 0.998 & 1.793 & 23.860 & 15.129 \end{array}\right]$$

Using Result 6.2.4 we can calculate $\hat{\rho}_{jl}|(q+1,\ldots,k)$ as

$$\frac{|\mathbf{S}_{11.2}|_{j,l}}{|\mathbf{S}_{11.2}|_{j,j}^{1/2}|\mathbf{S}_{11.2}|_{l,l}^{1/2}}$$

where

```
S_11 <- as.matrix(cov_mat[1:2,1:2])
S_12 <- as.matrix(cov_mat[1:2,3:4])
S_21 <- as.matrix(cov_mat[3:4,1:2])
S_22 <- as.matrix(cov_mat[3:4,3:4])

S_11
```

```
##        X_0    X_3
## X_0 60.516 21.122
## X_3 21.122 18.225
```

```
S_12
```

```
##       X_2   X_1
## X_0 3.511 0.998
## X_3 3.633 1.793
```

```
S_21
```

```
##       X_0   X_3
## X_2 3.511 3.633
## X_1 0.998 1.793
```

```
S_22
```

```
##        X_2    X_1
## X_2 54.756 23.860
## X_1 23.860 15.129
```

such that

```r
S_11.2 <- S_11 - S_12 %*% solve(S_22) %*% S_21
S_11.2
```

```
##           X_0      X_3
## X_0 60.23108 20.91265
## X_3 20.91265 17.97464
```

and

```r
rho_hat <- S_11.2[1,2] %*% 1/(S_11.2[1,1]^(1/2) %*% S_11.2[2,2]^(1/2))
rho_hat
```

```
##           [,1]
## [1,] 0.6355775
```

Problem 6.7

Suppose $\mathbf{S}$ is partitioned as in (6.2.9). Evaluate $\hat{\rho}^2_{0(1,\ldots,k)}$ in terms of $|\mathbf{S}|$, $|\mathbf{S}^{(1)}|$ and S_{00}.

From (6.2.9), we have that $\mathbf{S}$ is partitioned as

$$\mathbf{S} = \begin{pmatrix} S_{00} & \mathbf{s}_{01} \\ \mathbf{s}_{10} & \mathbf{S}^{(1)} \end{pmatrix}$$

By Result 6.2.4, we can estimate the multiple correlation coefficient $\hat{\rho}_{0(1,\ldots,k)}$ as

$$\frac{\left(\mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}\right)^{1/2}}{S_{00}^{1/2}}$$

Let's solve for $\hat{\rho}^2_{0(1,\ldots,k)}$.

$$\begin{aligned}
\hat{\rho}^2_{0(1,\ldots,k)} &= \left[\frac{\left(\mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}\right)^{1/2}}{S_{00}^{1/2}}\right]^2 \\
&= \frac{\mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}}{S_{00}} \\
&= \frac{S_{00} - S_{00} + \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}}{S_{00}} \\
&= \frac{S_{00} - \left[S_{00} - \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}\right]}{S_{00}} \\
&= \frac{S_{00} - |\mathbf{S}^{(1)}|^{-1}|\mathbf{S}^{(1)}|\left[S_{00} - \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}\right]}{S_{00}}
\end{aligned}$$

Since $S_{00} - \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}$ is a 1×1 matrix, it is its determinant, i.e., $(S_{00} - \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}) = |S_{00} -$

$\mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}|$ such that we have

$$\begin{aligned}
\hat{\rho}^2_{0(1,\ldots,k)} &= \frac{S_{00} - |\mathbf{S}^{(1)}|^{-1}|\mathbf{S}^{(1)}|\left[S_{00} - \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}\right]}{S_{00}} \\
&= \frac{S_{00} - |\mathbf{S}^{(1)}|^{-1}|\mathbf{S}^{(1)}|\left|S_{00} - \mathbf{s}_{01}[\mathbf{S}^{(1)}]^{-1}\mathbf{s}_{10}\right|}{S_{00}} \\
&= \frac{S_{00} - |\mathbf{S}^{(1)}|^{-1}|\mathbf{S}|}{S_{00}} \\
&= 1 - \frac{|\mathbf{S}^{(1)}|^{-1}|\mathbf{S}|}{S_{00}} \\
&= 1 - \frac{|\mathbf{S}|}{S_{00}|\mathbf{S}^{(1)}|}
\end{aligned}$$

Problem 6.8

When $\rho_{0(1,\ldots,k)} = 0$, show that the statistic $\hat{\rho}^2_{0(1,\ldots,k)}$ follow a Beta distribution. What are the parameters of the distribution?

Let $\rho_{0(1,\ldots,k)} = 0$. Then, from Result 6.2.5 we have that

$$\frac{(N-k-1)\hat{\rho}^2_{0(1,\ldots,k)}}{k(1-\hat{\rho}^2_{0(1,\ldots,k)})} \sim F_{(k,N-k-1)}$$

From Appendix 14, we know that if $V \sim F_{p,q}$ then

$$X = \frac{(p/q)V}{1+(p/q)V} \sim Beta(\frac{p}{2}, \frac{q}{2})$$

so let's consider $\frac{p}{q}V$, where $V = \hat{\rho}^2_{0(1,\ldots,k)}$, $p = k$, and $q = N-k-1$, such that we have

$$\frac{p}{q}V = \frac{k}{N-k-1}\left(\frac{(N-k-1)\hat{\rho}^2_{0(1,\ldots,k)}}{k(1-\hat{\rho}^2_{0(1,\ldots,k)})}\right) = \frac{\hat{\rho}^2_{0(1,\ldots,k)})}{1-\hat{\rho}^2_{0(1,\ldots,k)})}$$

Using Appendix 14, we can calculate X as

$$\begin{aligned}
X &= \frac{\frac{\hat{\rho}^2_{0(1,\ldots,k)}}{1-\hat{\rho}^2_{0(1,\ldots,k)}}}{1+\frac{\hat{\rho}^2_{0(1,\ldots,k)}}{1-\hat{\rho}^2_{0(1,\ldots,k)}}} \\
&= \frac{\frac{\hat{\rho}^2_{0(1,\ldots,k)}}{1-\hat{\rho}^2_{0(1,\ldots,k)}}}{\frac{1-\hat{\rho}^2_{0(1,\ldots,k)}+\hat{\rho}^2_{0(1,\ldots,k)}}{1-\hat{\rho}^2_{0(1,\ldots,k)}}} \\
&= \frac{\frac{\hat{\rho}^2_{0(1,\ldots,k)}}{1-\hat{\rho}^2_{0(1,\ldots,k)}}}{\frac{1}{1-\hat{\rho}^2_{0(1,\ldots,k)}}} \\
&= \frac{\hat{\rho}^2_{0(1,\ldots,k)}(1-\hat{\rho}^2_{0(1,\ldots,k)})}{1-\hat{\rho}^2_{0(1,\ldots,k)}} \\
&= \hat{\rho}^2_{0(1,\ldots,k)}
\end{aligned}$$

giving us that $X = \hat{\rho}^2_{0(1,\ldots,k)} \sim Beta(\frac{p}{2}, \frac{q}{2}) = Beta(\frac{k}{2}, \frac{N-k-1}{2})$.