Methods of Economic Investigation II (EC403)
Solutions to Problem Set #1

Oliver Linton

October 18, 2001

1. The non-zero condition is irrelevant. The (i, j)'th element of $X'X$ is $\sum_k x_{ik}x_{kj}$ and (j, i)'th element is $\sum_k x_{jk}x_{ki}$; these are the same for all i, j. Also, for any $a'(X'X)a = z'z = \sum_i z_i^2$, where the vector $z = Xa$. Clearly, $\sum_i z_i^2 \geq 0$ always. The only case when $a'(X'X)a = 0$ is if every element of the vector z is zero, i.e., $Xa = 0$. But if X is of full rank, $Xa = 0$ if and only if $a = 0$. Therefore if X is full rank then $X'X$ is strictly positive definite.

2. First note that $(AB)' = B'A'$ for any matrices A, B. If A is square and non-singular, then A' is also square and nonsingular so its inverse exists. We have

$$AA^{-1} = I = I' = (AA^{-1})' = (A^{-1})'A'.$$

   Since A and B are square, non-singular matrices, their inverses A^{-1} and B^{-1} exist. Then

$$AB(B^{-1}A^{-1}) = AIA^{-1} = I = (B^{-1}A^{-1})AB,$$

   which is the defining property of an inverse.

3. A is positive definite wheneever $x'Ax > 0$ for all $x \neq 0$. This implies that $Ax \neq 0$ for aevery $x \neq 0$, which implies that $Ax = 0$ only for $x = 0$, i.e., A is non-singular.

4. Let $y = a'x + b$, then $E(y) = a'\mu + b$ and $\text{var}(y) = a'\Sigma a$. In fact, $y \sim N(a'\mu + b, a'\Sigma a)$. The proof is based on moment generating function argument. If the moment generating function exists it uniquely characterizes the distribution [alternatively one can use characteristic functions which always exist, but are complex-valued in general]. The mgf of x is

$$E\exp(s'x) = \exp(s'\mu + s'\Sigma s/2).$$

   The mgf of y is

$$E\exp(ty) = E\exp(t(a'x + b)) = \exp(tb)E\exp(ta'x)$$

$$= \exp(tb)\exp(ta'\mu + t^2 a'\Sigma a/2)$$

$$= \exp(t(a'\mu + b) + t^2 a'\Sigma a/2)$$

by identifying $s = ta$. This is the mgf of a $N(a'\mu + b, a'\Sigma a)$ random variable.

5. Since A is symmetric and positive definite, there exists a representation

$$A = Q\Lambda Q',$$

where Λ is a diagonal matrix containing the eigenvalues of A and Q is an orthonormal matrix with $QQ' = I$ containing the eigenvectors of A. Define

$$P = A^{1/2} = Q\Lambda^{1/2}Q',$$

where $\Lambda^{1/2}$ is a diagonal matrix whose elements are the square roots of the corresponding Λ matrix. Then, P is symmetric and

$$PP' = Q\Lambda^{1/2}Q'Q\Lambda^{1/2}Q' = Q\Lambda^{1/2}I\Lambda^{1/2}Q' = Q\Lambda Q' = A.$$

In fact, P is not unique, because $P = Q\Lambda^{1/2}$ also satisfies $PP' = A$; however, $P = Q\Lambda^{1/2}$ is not a symmetrric matrix. Any covariance matrix is real and symmetric, and so is its inverse Σ^{-1}. Therefore, we can define the matrix $\Sigma^{-1/2}$ as above. Then

$$x'\Sigma^{-1}x = x'\Sigma^{-1/2}\Sigma^{-1/2}x = y'y = \sum_{i=1}^{n} y_i^2,$$

where $y = \Sigma^{-1/2}x$ [the matrix $\Sigma^{-1/2}$ is real and symmetric the first way we defined it]. The distribution of y is $N(0, I)$ so that we can conclude that $y'y$ is χ^2 with n degrees of freedom.

6. We can write $y = \sum\sum_{i,j} x_i x_j A_{ij}$. Therefore,

$$\frac{\partial^2 y}{\partial x_i \partial x_j} = A_{ij} + A_{ji},$$

so that the Hessian matrix is $A + A'$, which is equal to $2A$ in the symmetric case.

7. Let f_{XY} , $f_{X|Y}$, and f_X, f_Y be the joint, conditional and marginal densities of X and Y [we just treat the continuous case, but the discrete case is the same]. Then

$$E(X|Y = y) = \int x f_{X|Y}(x|y)dx = \int x \frac{f_{XY}(x, y)}{f_Y(y)} dx$$

by the definition of conditional density. Furthermore, by independence $f_{XY}(x,y) = f_X(x)f_Y(y)$ so that

$$\int x \frac{f_{XY}(x,y)}{f_Y(y)} dx = \int x \frac{f_X(x)f_Y(y)}{f_Y(y)} dx = \int x f_X(x) dx = E(X).$$

The reverse is not true. Is the reverse true? No. Suppose that

$$X = \varepsilon \sigma(Y),$$

where ε is independent of Y and has mean zero and variance one, while $\sigma(Y)$ is some non-trivial function. Then

$$E(X|Y) = E(\varepsilon\sigma(Y)) = \sigma(Y)E(\varepsilon|Y) = 0 = E(X).$$

However,

$$\text{var}(X|Y) = \text{var}(\varepsilon|Y)\sigma^2(Y) = \sigma^2(Y),$$

which does vary with Y. Thus, X is not independent of Y. To be specific, suppose that ε is standard normal and that $\sigma(Y) = Y$, where $Y = 1$ with probability $1/2$ and 2 with proobability $1/2$. Then

$$f_{X|Y}(x|1) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \text{ and } f_{X|Y}(x|2) = \frac{1}{2\sqrt{2\pi}} \exp(-x^2/8),$$

while

$$f_X(x) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} \exp(-x^2/2) + \frac{1}{2}\frac{1}{2\sqrt{2\pi}} \exp(-x^2/8).$$

Therefore, $f_{X|Y} \neq f_X$.

8. We have

$$\begin{aligned} \text{cov}(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] = E(XY) - E(X)E(Y) \\ &= E(E(X|Y)Y) - E(X)E(Y) \\ &= E(X)E(Y) - E(X)E(Y) = 0. \end{aligned}$$

The reverse is not true in general. Suppose that

$$X = Y^2,$$

where $Y \sim N(0,1)$. Then

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y) = E(Y^3) - E(Y^2)E(Y) = 0.$$

However,

$$E(X|Y) = Y^2 \neq 0$$

except when $Y = 0$.

9. We have

$$
\begin{aligned}
E\widehat{\beta} &= \beta \quad ; \quad \text{var}(\widehat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \\
E\widetilde{\beta} &= \beta \quad ; \quad \text{var}(\widetilde{\beta}) = \frac{n\sigma^2}{\left(\sum_{i=1}^n x_i\right)^2} \\
E\overline{\beta} &= \beta \quad ; \quad \text{var}(\overline{\beta}) = \sigma^2 \frac{1}{n^2} \sum_{i=1}^n \frac{1}{x_i^2}.
\end{aligned}
$$

By the Gauss-Markov theorem we know that $\widehat{\beta}$ is the best linear unbiased estimator. Therefore, we know that it beats $\widetilde{\beta}$ and $\overline{\beta}$ according to variance. This can be verified directly as well. But what about the comparison between $\widetilde{\beta}$ and $\overline{\beta}$? There is no uniform ranking, it just depends on the sample. For example, if $n = 2$, $x_1 = 1$, and $x_2 = 0$, $\text{var}(\overline{\beta}) = \infty$, while $\text{var}(\widetilde{\beta}) = 2\sigma^2$, so that $\widetilde{\beta}$ is better in this case. On the other hand, if $x_1 = 1$, and $x_2 = -1$, then $\text{var}(\widetilde{\beta}) = \infty$, while $\text{var}(\overline{\beta}) = \sigma^2/2$, so that $\overline{\beta}$ is better in this case.

10. The iterated least squares calculation should converge to $\widehat{\alpha}, \widehat{\beta}$. This is really a version of the so-called Jacobi or Gauss-Seidel iterative method that is used for solving large linear systems.