James K. Pringle
140.751
Dr. Brian Caffo
Assignment 2
23 October 2012, Tuesday

Problems 2.1, 2.11, 3.7

(2.1) Let X be a multivariate vector with mean μ. Show that $E[AX+b] = A\mu + b$.

Proof: Let X and b be vectors of length n. Let A be a matrix with dimensions $m \times n$. Thus $AX+b$ is a vector of length n. The i-th element of $AX+b$ is

$$\left(\sum_{j=1}^{n} a_{ij}x_j\right) + b_i$$

where a_{ij} is the element in the i-th row and j-th column of A and the indexed elements of X and b, x_j and b_j, respectively, are the corresponding components of those vectors. The expected value of an array is just the componentwise expected value. Hence, the expected value of the i-th element of $AX+b$ is

$$\begin{aligned} E[AX+b]_i &= E[(\sum_{j=1}^{n} a_{ij}x_j) + b_i] \\ &= (\sum_{j=1}^{n} a_{ij}E[x_j]) + b_i \text{ by the linearity of expected value} \\ &= (\sum_{j=1}^{n} a_{ij}\mu_i) + b_i \ . \end{aligned}$$

This is precisely the same as the i-th element of $A\mu + b$. Thus, $E[AX+b] = A\mu + b$.

(2.11) Let $X \sim N(0, I)$. Argue that if $AA' = I$, then $AX \sim N(0, I)$. Argue geometrically why this occurs.

Proof: From page 6 of Brian's scanned notes, we have the following.

Let $y = a + \omega X$ when $X \sim (\mu, \Sigma)$. Then $\mathrm{var}(y) = \omega\Sigma\omega'$.

Thus $\mathrm{var}(AX) = AIA' = AA' = I$. From (2.1), we have $E[AX] = A\mu = 0$. Thus $AX \sim N(0, I)$. This occurs because A is an orthogonal matrix by definition. Thus AX is merely a rotation of X that does not scale X. Therefore we would expect AX to have the same multi-variate distribution as X.

(3.7) Consider a linear model with iid $N(0, \sigma^2)$ errors. Show that $\frac{1}{n-p}e'e$, where e is the vector of residuals, is the ML estimate of σ^2. Further show that this estimate is unbiased.

Proof: Assume e is a vector of n errors and p is the rank and number of columns of X with $n > p$. Since the errors are iid, their joint density function is $\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\frac{x^2}{\sigma^2}\}$. The log of this density is

$$\log(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\frac{x^2}{\sigma^2}\}) = -\frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{n} -\frac{1}{2}\frac{x_i^2}{\sigma^2}$$

Now we take the derivative with respect to σ^2 because we want to maximize this expression with respect to σ^2. Setting the derivative equal to zero, we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial \sigma^2}\left(-\frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{n} -\frac{1}{2}\frac{x_i^2}{\sigma^2}\right) \\
&= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma^4}
\end{aligned}$$

Now multiply by σ^4 and solve for σ^2. We have

$$\begin{aligned}
0 &= -\frac{n}{2}\sigma^2 + \frac{1}{2}\sum_{i=1}^{n} x_i^2 \\
n\sigma^2 &= \sum_{i=1}^{n} x_i^2 \\
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\
\sigma^2 &= \frac{1}{n} e'e \ .
\end{aligned}$$

This value of σ^2 maximizes the log of the joint density function because the second derivative evaluated at this value is negative.

Notice $e = Y - X(X'X)^{-1}X'Y = (I_n - P)Y$. To show the above value, $\frac{1}{n-p}e'e$, is unbiased, we take the expected value of it.

$$\begin{aligned}
E[\frac{1}{n-p}e'e] &= E[\frac{1}{n-p}Y'(I_n - P)'(I_n - P)Y] \\
&= \frac{1}{n-p}E[Y'(I_n - P)Y] \text{ since } I_n - P \text{ is idempotent and symmetric} \\
&= \frac{1}{n-p}\mathrm{tr}((I_n - P)\mathrm{Var}(Y)) + E[Y]'(I_n - P)E[Y] \text{ by expected value of quadratic forms} \\
&= \frac{1}{n-p}\mathrm{tr}((I_n - P)\sigma^2 I) \text{ since } E(Y) = 0 \\
&= \frac{1}{n-p}\sigma^2(n-p) \\
&= \sigma^2
\end{aligned}$$

since $\mathrm{tr}(P) = \mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}(X'X(X'X)^{-1}) = \mathrm{tr}(I_p) = p$. Hence we have that $\frac{1}{n-p}e'e$ is an unbiased estimator of the variance of the errors.

For the rest of this problem, we use the following result from Seber and Lee.

> If $y \sim N(0, \Sigma)$ then $y'Ay \sim \chi^2$ with $\mathrm{rank}(A\Sigma)$ degrees of freedom if $A\Sigma$ is idempotent.

(a) Argue that $\frac{1}{\sigma^2}(y - X\beta)'(y - X\beta)$ is χ^2_n.

Proof: Rewrite the expression as $(y - X\beta)'\frac{1}{\sigma^2}I_n(y - X\beta)$. From the discussion above, $y - X\beta \sim N(0, \sigma^2 I_n)$. Notice that $\frac{1}{\sigma^2}I_n\sigma^2 I_n = I_n$ is idempotent and has rank n. Thus, $(y - X\beta)'\frac{1}{\sigma^2}I_n(y - X\beta) \sim \chi^2_n$.

(b) Argue that $\frac{1}{\sigma^2}e'e$ is χ^2_{n-p}.

Proof: This is the start of an incorrect proof. As above, notice that $\frac{1}{\sigma^2}e'e = Y'\frac{1}{\sigma^2}(I_n - P)Y$. The random vector Y has variance $\sigma^2 I_n$. The matrix $\frac{1}{\sigma^2}(I_n - P)\sigma^2 I_n = I_n - P$ is idempotent and has rank equal to $\mathrm{rank}(I_n - P) = \mathrm{tr}(I_n - P) = \mathrm{tr}(I_n) - \mathrm{tr}(P) = n - p$ (see Seber and Lee, page 28).

Therefore $\frac{1}{\sigma^2}e'e$ is χ^2_{n-p}. I think Brian might have been going for something like that. But that is incorrect because $Y \not\sim N(0, \Sigma)$ since $E[Y]$ is not necessarily 0.

On the other hand, $e = y - X\beta$. Therefore $\frac{1}{\sigma^2}e'e = \frac{1}{\sigma^2}(y - X\beta)'(y - X\beta)$. So problem (b) is the exact same as (a).

(c) Argue that $\frac{1}{\sigma^2}(y - X\beta)'X(X'X)^{-1}X'(y - X\beta)$ is χ^2_p.

Proof: As in (a), $y - X\beta \sim N(0, \sigma^2 I_n)$. Calculating, we see, $\frac{1}{\sigma^2}X(X'X)^{-1}X'\sigma^2 I_n = P$, which is idempotent. The rank of P is p since each of X, $(X'X)^{-1}$, and X' has rank p. Therefore, $\frac{1}{\sigma^2}(y - X\beta)'X(X'X)^{-1}X'(y - X\beta)$ is χ^2_p.

(d) The expected value of a quadratic form is

$$E[Y'AY] = \mathrm{tr}(A\Sigma) + \mu' A \mu$$

where μ and Σ are the expected value and covariance matrix, respectively of Y. In the three problems above, taking the corrected form of (b), $\mu = 0$. Hence we are left with $E[Y'AY] = \mathrm{tr}(A\Sigma)$. It is clear that in (a), (b), and (c), $\mathrm{tr}(A\Sigma)$ is the same as the rank of $A\Sigma$, which is the degrees of freedom of the χ^2 distribution according to the theorem. In (a) and (b), $A\Sigma = I_n$, and it is obvious that $\mathrm{tr}(I_n) = \mathrm{rank}(I_n) = n$. In (c), $A\Sigma = P$. An argument is given above in two different spots why $\mathrm{tr}(P) = \mathrm{rank}(P) = p$.