Sample Problems for Stat 750 In-Class Test

Eric Slud March 12, 2022

1. (*Proof of independence of Wishart quadratic forms with orthogonal matrices*) Prove in detail that if Y is an $n \times p$ normal data matrix with mean $\mathbf{0}$ and variance Σ for each row, and if B and C are nonnegative-definite symmetric matrices such that $BC = 0$, then $Y'BY$ and $Y'CY$ are independent, and if B and C are also projection matrices then these two quadratic forms are Wishart random $p \times p$ matrices, and give the Wishart parameters.

2. (*Simulation*) Let Y be an 100×5 matrix with independent identically distributed rows

$$\mathbf{Y}_i = \mu + V_i \cdot \Sigma^{1/2} \mathbf{U}_i \quad , \quad V_i \sim g(v) = 8\,(1+v)^{-9} \quad \text{for } v > 0 \,, \quad \mathbf{U}_i \sim \text{unif. unit vector}$$

where $\mu \in \mathbb{R}^5$ and the 5×5 nonsingular covariance matrix Σ are unknown, and V_i and $\mathbf{U}_i$ are independent for each i. Explain as clearly as you can the steps by which you would simulate $R = 1e4$ replications of 100×5 datasets and estimate the 0.9, 0.95 quantiles of the Hotelling T^2 test statistic for the data Y that would be used to test the hypothesis $\mathbf{H}_0 : \mu = \mathbf{0}$.

3. (*Multivariate Normal - best linear approximation – application to cross-moments*) Suppose that $X = (X_1, X_2, X_3)$ is a normal random vector with mean 0 and covariance matrix

$$\begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}$$

Find $E(X_1 \cdot X_2^2 \cdot X_3)$.

4. (*Spherical symmetry*) (a). Show that if $\mathbf{X} = (X_1, \ldots, X_k)$ is spherically symmetric, for $k \geq 3$, then $(X_1, \ldots, X_{k-1})$ is also spherically symmetric. (b) Show that if if $\mathbf{X} = (X_1, \ldots, X_k)$ is spherically symmetric, and the positive scalar random variable V is independent of $\mathbf{X}$, then $V \cdot \mathbf{X}$ is also spherically symmetric.

5. (*Recognition of distribution of various test statistics*) Suppose Y is an $n \times 4$ normal data matrix with rows $\sim \mathcal{N}_4(\mu, \Sigma)$. Let $\hat{\mu}, S$ be the MLEs for μ, Σ respectively.

(a). Let

$$A = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{pmatrix}$$

What is the distribution of $(A\hat{\mu})' \, (ASA')^{-1} \, (A\hat{\mu})$? This expression has been corrected. For what hypothesis can it be used as a test statistic ?

(b) What does $n^{-1} \, Y'(I - n^{-1} \underline{1}^{\otimes 2}) \, Y$ estimate, and what is its distribution, where I is the $n \times n$ identity matrix and $\underline{1}$ the n-vector of all 1's ?

6. In the setting of problem 5, find the Union Intersection Test for $\mathbf{H}_0 : \ \mu_1 + \mu_2 - \mu_3 = a, \ \mu_1 - \mu_2 + \mu_4 = b$, with Σ unknown and unrestricted and a, b fixed. Use this test to provide simultaneous confidence intervals for $\mu_1 + \mu_2 - \mu_3$ and $\mu_1 - \mu_2 + \mu_4$.

7. (*Reasoning with projections and Wisharts*) Suppose that $Y = (X \,|\, Z))$ is an $n \times p$ normal data matrix with mean $\mathbf{0}$ and variance Σ for each row, where X and Z are respectively normal data matrices with q and $r = p - q$ columns, with Σ partitioned into blocks as

$$\Sigma \; = \; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad , \qquad \Sigma_{11} = \mathrm{Var}(\mathbf{X}_i) \quad \text{invertible}$$

Let

$$M \; = \; Y'Y \; = \; \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \quad , \qquad M_{11} = X'X \;\; , \quad M_{22} = Z'Z \;\; , \quad M_{12} = X'Z$$

Show that conditionally given Z, the random matrix $M_{11} - M_{12} M_{22}^{-1} M_{21} \equiv X'X - X'Z(Z'Z)^{-1} Z'X$ has the form $X'QX$ for an $n \times n$ projection matrix Q, and therefore has a Wishart distribution independent of Z; and find the parameters of that Wishart distribution.

8. (*Max-Min problem for Matrices*) Suppose that B is a fixed known symmetric positive-definite $p \times p$ matrix. Find the supremum over symmetric nonnegative definite matrices A of $(\det(A))^{1/p} - \mathrm{trace}(A \, B^{-1})$.

9. (*Regression*) Consider the following results concerning a simulated dataset SimProbFr, a data-frame consisting of $p = 2$ outcome column Y1, Y2 with 3 grouping factors F1 (2 levels H, L), F2 (3 levels A,B,C) and F3 (3 levels 1,2,3). Altogether there are 18 group-combinations, and 20 observations for each combination, for a total of $n = 360$ observations.

```
> SimProbFr[20*(1:18),]
    F1 F2 F3          Y1         Y2
20   H  A  1  5.25103606  6.7797895
40   H  A  2  2.06371453  2.1915714
60   H  A  3  1.94659562  4.4645685
80   H  B  1  0.12130146  5.1505129
100  H  B  2  1.84702204 -0.3419188
120  H  B  3 -0.34130798 -2.1825328
140  H  C  1  0.05014656  2.6975613
160  H  C  2 -3.47980201  5.3432956
180  H  C  3  1.32086595 -0.6060656
200  L  A  1  5.74936203  5.8598511
220  L  A  2  0.56104101  1.0059033
240  L  A  3  7.34881370  8.6745729
260  L  B  1 -4.45722390 -4.3722358
280  L  B  2 -0.51043199  3.2898355
300  L  B  3  0.02937361 -2.5618660
320  L  C  1  0.12753017 -0.8846669
340  L  C  2 -1.49726159  0.3942742
360  L  C  3  2.92435096  6.1544452
```

(a). Here are three separate model-fits and some results:

```
> fit1 = lm(cbind(Y1,Y2) ~ F1, data=SimProbFr)
> t(fit1$eff) %*% fit1$eff
         Y1       Y2
Y1 3528.838 2342.430
Y2 2342.430 5003.926
> t(fit1$res) %*% fit1$res
         Y1       Y2
Y1 3318.471 1689.587
Y2 1689.587 2973.985
```

```
> fit2 = lm(cbind(Y1,Y2) ~ F2, data=SimProbFr)
> t(fit2$eff) %*% fit2$eff
         Y1       Y2
Y1 3528.838 2342.430
Y2 2342.430 5003.926
> t(fit2$res) %*% fit2$res
         Y1       Y2
Y1 2870.282 1281.629
Y2 1281.629 2623.436

> fit3 = lm(cbind(Y1,Y2) ~ F3, data=SimProbFr)
> t(fit3$eff) %*% fit3$eff
         Y1       Y2
Y1 3528.838 2342.430
Y2 2342.430 5003.926
> t(fit3$res) %*% fit3$res
         Y1       Y2
Y1 3299.505 1659.491
Y2 1659.491 2931.040
```

(a). Is there enough information in just these displayed matrices to calculate Wilks statistics separately for the MANOVA for the three factor-groupings ? Explain.

(b). I give you next a series of MANOVA tables showing you the Wilks Lambda statistic, and the problem will be to interpretwhat hypotheses are being tested in each table and to interpret the results.

```
> anova(fit1, test="Wilks")
Analysis of Variance Table
             Df   Wilks approx F num Df den Df Pr(>F)
(Intercept)   1 0.58057  128.954      2    357 <2e-16 ***
F1            1 0.98755    2.251      2    357 0.1068
Residuals   358
```

```
>   anova(fit2, test="Wilks")
Analysis of Variance Table
             Df   Wilks approx F num Df den Df    Pr(>F)
(Intercept)   1 0.55999  139.861      2    356 < 2.2e-16 ***
F2            2 0.82889   17.512      4    712 1.021e-13 ***
Residuals   357

>  anova(fit3, test="Wilks")
Analysis of Variance Table
             Df   Wilks approx F num Df den Df  Pr(>F)
(Intercept)   1 0.57755  130.197      2    356 < 2e-16 ***
F3            2 0.97385    2.374      4    712 0.05084 .
Residuals   357
```