Ec240a, Fall 2023

Professor Bryan Graham

Problem Set 1

Due: October 27th, 2023

Problem sets are due at 5PM. The GSI will provide instructions on how to turn in your problem set. You may work in groups, but each student should turn in their own write-up (including a "printout" of a narrated/commented and executed Jupyter Notebook if applicable). Please also e-mail a copy of any Jupyter Notebook to the GSI (if applicable).

1 Multivariate normal distribution

Let $\mathbf{Y} = (Y_1, \ldots, Y_K)'$ be a $K \times 1$ random vector with density function

$$f(y_1, \ldots, y_K) = (2\pi)^{-K/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)' \Sigma^{-1} (\mathbf{y} - \mu)\right),$$

for Σ a symmetric positive definite $K \times K$ matrix and μ a $K \times 1$ vector. We say that $\mathbf{Y}$ is a multivariate normal random variable with mean μ and covariance Σ or

$$\mathbf{Y} \sim \mathcal{N}(\mu, \Sigma).$$

The multivariate normal distribution plays an outsized role in econometrics. A mastery of its basic properties is essential for both applied and theoretical work in economics. The distribution first arose in the second edition of Abraham de Moivre's *The Doctrine of Chances* in the context of approximating binomial probabilities (de Moivre essentially proved a special case of the central limit theorem). It appears again in Gauss' 1809 *Theoria Modus Corporum Coelestium* book as an "error" distribution for astronomical measurements. In this book Gauss' uses the normal distribution to motivate the method of least squares, which he used to correctly predict the future location of the dwarf planet *Ceres*.

Due to its relationship with the central limit theorem (CLT), the normal distribution is an indispensable tool for all econometricians (applied or theoretical). This problem provides an opportunity for you to derive some key properties of the normal distribution. *You should commit these properties to memory.* You may find Goldberger (1991) and Anderson (2005), among other references, useful in showing these results. Most of what follows is also discussed on Wikipedia, Stack Exchange and so on.

1. For
$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$
with $\Sigma_{12} = \Sigma_{21}'$ show that
$$\mathbf{Y}_1 \sim \mathcal{N}(\mu_1, \Sigma_{11}).$$

2. Likewise show that
$$\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1 \sim \mathcal{N}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).$$
(you may need to review partitioned matrix inverses)

3. If $\Sigma_{12} = 0$, show that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent.

4. Let $\mathbf{Z} = \mathbf{A} + \mathbf{B}\mathbf{Y}$ for $\mathbf{A}$ and $\mathbf{B}$ non-random and $\mathbf{B}$ with full row rank, show that

$$\mathbf{Z} \sim N\left(\mathbf{A} + \mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}'\right).$$

5. Set $\mathbf{A} = 0$ and $\mathbf{B} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Show that for $\mathbf{Z} = \mathbf{A} + \mathbf{B}\mathbf{Y} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ that

   (a) Z_1 and Z_2 are univariate normal random variables;

   (b) their joint distribution is *not* bivariate normal;

   (c) Explain [2 to 3 sentences].

6. Let $\{\mathbf{Y}_i\}_{i=1}^{N}$ be a random sample of size N drawn from the multivariate normal population described above. Show that $\sqrt{N}\left(\overline{\mathbf{Y}} - \mu\right)$ is a $\mathcal{N}(0, \Sigma)$ random variable for $\overline{\mathbf{Y}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Y}_i$, the sample mean (HINT: Use independence of the $i = 1, \ldots, N$ draws and your result in Problem 4 above).

7. Let $W = (\mathbf{Y} - \mu)' \Sigma^{-1} (\mathbf{Y} - \mu)$. Show that $W \sim \chi^2_K$ (i.e., W is a chi-square random variable with K degrees of freedom).

8. Let $\mathbf{W} = N \cdot \left(\overline{\mathbf{Y}} - \mu\right)' \Sigma^{-1} \left(\overline{\mathbf{Y}} - \mu\right)$. Show that $\mathbf{W} \sim \chi^2_K$.

9. Let $\chi_K^{2,1-\alpha}$ be the $(1-\alpha)^{th}$ quantile of the χ^2_K distribution (i.e., the number satisfying the equality $\Pr\left(\mathbf{W} \leq \chi_K^{2,1-\alpha}\right) = 1 - \alpha$ with $\mathbf{W}$ a chi-square random variable with K degrees of freedom). Let D be a $P \times K$ $(P \leq K)$ matrix of rank P and d a $P \times 1$ vector of constants. Consider the hypothesis

$$H_0 : D\mu = d$$
$$H_1 : D\mu \neq d.$$

   Maintaining H_0 derive the sampling distribution of $D\overline{\mathbf{Y}}$ as well as that of

$$\mathbf{W} = N \cdot \left(D\overline{\mathbf{Y}} - d\right)' (D\Sigma D)^{-1} \left(D\overline{\mathbf{Y}} - d\right).$$

   You observe that, for the sample in hand, $\mathbf{W} > \chi_P^{2,1-\alpha}$ for $\alpha = 0.05$. Assuming H_0 is true, what is the ex ante (i.e., pre-sample) probability of this event? What are you inclined to conclude after observing $\mathbf{W}$ in the sample in hand?

2 Normal Bayesian Learning

Each morning you run home from work. You do this because regular aerobic exercise improves cognitive function and is enjoyable. Prior to your first such run, at time $t = 0$, you believe the actual distance of your commute, denoted by A, is about α_0 with about a 68% chance that is between $\alpha_0 - \tau$ and $\alpha_0 + \tau$. More specifically your subjective beliefs about the distance of your commute are describe by the $\mathcal{N}\left(\alpha_0, \tau^2\right)$ distribution. One way to think about your beliefs is that you behave "as if" the actual distance is the random draw

$$A \sim \mathcal{N}\left(\alpha_0, \tau^2\right).$$

Each period $t=1,\ldots,T$ you run home. Your GPS watch measures the distance of each such run as $Y_0,\ldots,Y_T$. After reading the technical documentation accompanying your watch you conclude that these measurements are also Gaussian random variables:

$$Y_t|\,A \overset{iid}{\sim} \mathcal{N}\left(A,\sigma^2\right),\ t=1,\ldots,T.$$

In what follows you may treat both τ^2 and σ^2 as known. Let $\mathbf{Y}_T=(Y_1,\ldots,Y_T)$.

1. Show that, unconditional on A,

$$\mathbf{Y}_T \overset{iid}{\sim} \mathcal{N}\left(\alpha_0\iota_T,\sigma^2 I_T+\tau^2\iota_T\iota_T'\right),$$

   where ι_T is a $T\times 1$ vector of ones and I_T the $T\times T$ identity matrix (<u>HINT:</u> Use your result from Problem 4 in Part 1 above).

2. Show that $\mathbb{C}\left(Y_t,A\right)=\tau^2$ and hence that

$$\left(\begin{array}{c} A \\ \mathbf{Y}_T \end{array}\right) \sim N\left(\left(\begin{array}{c} \alpha_0 \\ \alpha_0\iota_T \end{array}\right),\left(\begin{array}{cc} \tau^2 & \tau^2\iota_T' \\ \tau^2\iota_T & \sigma^2 I_T+\tau^2\iota_T\iota_T' \end{array}\right)\right).$$

3. Using this result show that

$$\mathbb{E}\left[A|\,\mathbf{Y}_T\right]=\frac{\sigma^2}{\sigma^2+T\tau^2}\alpha_0+\frac{T\tau^2}{\sigma^2+T\tau^2}\bar{Y}.$$

   You may find the matrix inverse results in Henderson & Searle (1981) helpful for deriving this expression (along with your results from Part 1 above).

4. Similarly show that

$$\mathbb{V}\left(A|\,\mathbf{Y}_T\right)=\frac{\sigma^2\tau^2}{\sigma^2+T\tau^2}$$

   and hence that

$$A|\,\mathbf{Y}_T\sim\mathcal{N}\left(\frac{\sigma^2}{\sigma^2+T\tau^2}\alpha_0+\frac{T\tau^2}{\sigma^2+T\tau^2}\bar{Y},\frac{\sigma^2\tau^2}{\sigma^2+T\tau^2}\right).$$

5. Describe your (posterior) beliefs about the distance of your compute, A, after completing T runs home [3-5 sentences].

6. Let $\bar{\sigma}^2_{A,t}=\mathbb{V}\left(A|\,\mathbf{Y}_t\right)$ denote your posterior variance for A after t runs home. Show that

$$\bar{\sigma}^2_{A,t}=\frac{1}{\frac{1}{\bar{\sigma}^2_{A,t-1}}+\frac{1}{\sigma^2}}.$$

   Similarly, let$\bar{\alpha}_t=\mathbb{E}\left[A|\,\mathbf{Y}_t\right]$ denote your posterior mean for A after t runs home. Show that

$$\bar{\alpha}_{t+1}=\bar{\alpha}_t+\frac{1/\sigma^2}{1/\sigma^2+1/\bar{\sigma}^2_{A,t}}\left(Y_{t+1}-\bar{\alpha}_t\right),$$

   and hence that you can update your beliefs recursively.

7. A veteran athlete in your neighborhood wheel measures the distance of your commute with arbitrary precision. This athlete ask you to form a best guess for the actual distance based on the information you have collected up to period T. Assume that your utility you get from guessing distance a is $-(A-a)^2$. You choose $\hat{\alpha}$ to maximize expected utility

$$\hat{\alpha} = \arg\max_a \mathbb{E}\left[-(A-a)^2 \middle| \mathbf{Y}_T \right].$$

Show that

$$\mathbb{E}\left[-(A-a)^2 \middle| \mathbf{Y}_T \right] = -(\bar{\alpha}_T - a)^2 - \bar{\sigma}_{A,t}^2$$

and hence that your optimal distance estimate is the posterior mean $\hat{\alpha} = \bar{\alpha}_T$. Discuss [3-6 sentences].

3 Further Reading

Stigler (1990) and Stigler (2002) provide engaging historical background on the development of the Normal Distribution, the Central Limit Theorem, and Bayes' Rule. Anderson (2005) collects important properties of the multivariate normal distribution (with proofs).

A useful matrix algebra reference for econometricians is Magnus (2019). Horn & Johnson (2013) provide a nice introduction to matrix analysis.

The simple normal learning model featured in Part 2 arises frequently in applied micro; see Foster & Rosenzweig (1995), Jovanovic & Nyarko (1996) and Altonji & Pierret (2001) for interesting examples and Chamley (2003) for a textbook discussion.

References

Altonji, J. G. & Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116(1), 313 – 350.

Anderson, T. W. (2005). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons, Inc., 3rd edition.

Chamley, C. P. (2003). *Rational Herds: Economic Models of Social Learning*. Cambridge: Cambridge University Press.

Foster, A. D. & Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, 103(6), 1176 – 1209.

Goldberger, A. S. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Henderson, H. V. & Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1), 53 – 60.

Horn, R. A. & Johnson, C. R. (2013). *Matrix Analysis*. Cambridge: Cambridge University Press, 2nd edition.

Jovanovic, B. & Nyarko, Y. (1996). Learning by doing and the choice of technology. *Econometrica*, 64(6), 1299 – 1310.

Magnus, J. R. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: John Wiley & Sons, Inc.

Stigler, S. M. (1990). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: Harvard University Press.

Stigler, S. M. (2002). *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press.