ECOM40006/90013 Econometrics 3

Week 2 Extras: Solutions

Question 1: Revisiting basic definitions

For this question, let's just explain the roles of each of these functions in turn.

- A distribution function is a function which describes a probability distribution over a set of possible events. That is: before talking about a probability distribution, one first needs to figure out the set that it's defined on.

  For example, if the set of possible events happens to be the real line (i.e. $\mathbb{R}$. In this sense, anything that we see will be a real number, heuristically speaking), then we are said to be dealing with probabilities on $\mathbb{R}$.

  If you wish to be abstract, then here is a very formal definition of a distribution function:

  Distribution function. The *distribution function* of a probability $\mathbf{P}$ on $\mathbb{R}$ is the function $F : \mathbb{R} \to \mathbb{R}$ defined by

  $$F(x) = \mathbf{P}((-\infty, x]), \qquad x \in \mathbb{R}.$$

  More heuristically, you'll usually refer to it as something much more familiar:

  $$F(x) = \mathbf{P}(X \leq x).$$

  A distribution function has the usual properties that we would expect from a cumulative distribution function. On the real line for example,

  - F is non-decreasing
  - F is right-continuous[1]
  - $\lim_{x \to \infty} F(x) = 1$
  - $\lim_{x \to -\infty} F(x) = 0$

[1] At any point, F takes the value of its 'right' limit. This is to account for jumps in the function, which is especially noticeable whenever you have anything that's discrete.

- A (probability) density function is defined to be the derivative of the CDF in the case where the probability distribution is (absolutely) continuous.[2] Formally, it can be written as such: if there exists a function $f : \mathbb{R} \to \mathbb{R}_+$ such that

$$\int_{-\infty}^{t} f(x)\, dx = F(t),$$

then $f(x)$ is said to be the probability density function of F. Note that this implies a result that we might be quite familiar with:

$$f(x) = \frac{dF(x)}{dx}.$$

The usual properties of a PDF hold: it is (i) non-negative and (ii) it integrates to 1.

Note that we don't interpret it as a probability in the usual sense: when dealing with PDFs, the probability of observing an exact point is zero. To see why, just consider this:

$$\mathbf{P}(X = a) = \mathbf{P}(a \leq X \leq a) = \int_{a}^{a} f(x)\, dx = 0.$$

- A probability (mass) function is used for discrete probabilities. For general purposes we can think of it as

$$p(x) = \mathbf{P}(X = x).$$

The usual properties hold: $p(x)$ is (i) non-negative and (ii) sums to 1.

Question 2: The bivariate normal distribution

(a) Recall that the determinant of a 2×2 matrix is

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Direct calculation yields:

$$\begin{aligned} |\Sigma|^{-1/2} &= \det \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}^{-1/2} \\ &= \frac{1}{\sqrt{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2}} \\ &= \frac{1}{\sigma_1 \sigma_2 \sqrt{\left(1 - \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}\right)}} \\ &= \frac{1}{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \end{aligned}$$

[2] In plain(er) English: it's for continuous random variables.

(b) Here, use the inverse of a 2×2 matrix. The formula is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

To simplify the notation, let's define

$$a = x_1 - \mu_1 \qquad \text{and} \qquad b = x_2 - \mu_2.$$

Then inside the brackets of the joint pdf we have (ignoring the $-1/2$)

$$\begin{aligned}
\begin{pmatrix} a & b \end{pmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}^{-1} \begin{pmatrix} a \\ b \end{pmatrix} &= \frac{1}{\det(\Sigma)} \begin{pmatrix} a & b \end{pmatrix} \begin{bmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{bmatrix}^{-1} \begin{pmatrix} a \\ b \end{pmatrix} \\
&= \frac{1}{\det(\Sigma)} \begin{bmatrix} a\sigma_2^2 - b\sigma_{12} & -a\sigma_{12} + b\sigma_1^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\
&= \frac{1}{\det(\Sigma)} (a^2\sigma_2^2 - ab\sigma_{12} - ab\sigma_{12} + b\sigma_1^2) \\
&= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} a^2\sigma_2^2 - 2ab\sigma_{12} + b\sigma_1^2) \\
\\
&= \frac{1}{1-\rho^2} \left[\frac{a^2\sigma_2^2}{\sigma_1^2\sigma_2^2} + \frac{b^2\sigma_1^2}{\sigma_1^2\sigma_2^2} - \frac{2ab\sigma_{12}}{\sigma_1^2\sigma_2^2} \right] \\
&= \frac{1}{1-\rho^2} \left[\frac{a^2}{\sigma_1^2} + \frac{b^2}{\sigma_2^2} - 2\frac{\sigma_{12}}{\sigma_1\sigma_2} \frac{ab}{\sigma_1\sigma_2} \right]
\end{aligned}$$

Once we expand out a and b, we obtain the answer we were looking for:

$$\begin{pmatrix} a & b \end{pmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}^{-1} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{1-\rho^2} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right]$$

(c) Before doing anything further, first let

$$z_1 = \frac{x_1 - \mu_1}{\sigma_1} \qquad \text{and} \qquad z_2 = \frac{x_2 - \mu_2}{\sigma_2}$$

so that

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} (z_1^2 + z_2^2 - 2\rho z_1 z_2) \right).$$

Furthermore, let's use the final answer to give us an indicator of what to do in our working.[3] Observe that

$$\begin{aligned}
\Bigg(\underbrace{[x_2 - \mu_2]}_{a} - \underbrace{\left[\rho \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \right]}_{b} \Bigg)^2 &= a^2 - 2ab - b^2 \\
&= (x_2-\mu_2)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)(x_2-\mu_2) + \rho^2 \frac{\sigma_2^2}{\sigma_1^2}(x_1-\mu_1)^2. \quad (1)
\end{aligned}$$

[3] It's either this or we ask ourselves how confident we are in factorizing a quadratic with a gratuitous amount of Greek.

Alright, so when we do our calculations, the numerator in our $\exp(\cdot)$ term will need to look like this. So here goes: by the definition of a conditional density,

$$\begin{aligned} f(x_2|x_1) &= \frac{f(x_1, x_2)}{f(x_1)} \\ &= \frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\dfrac{1}{2(1-\rho^2)}(z_1^2 + z_2^2 - 2\rho z_1 z_2)\right)}{\dfrac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\dfrac{1}{2}z_1^2\right)} \end{aligned}$$

Algebraic manipulation on the fractions on the left and power laws for the $\exp(\cdot)$ terms on the right gives

$$f(x_2|x_1) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 + z_2^2 - 2\rho z_1 z_2) + \frac{1}{2}z_1^2.\right)$$

Let's focus specifically on the contents of the $\exp(\cdot)$ term for now. Observe first that

$$\frac{1}{2}z_1^2 \times \frac{1-\rho^2}{1-\rho^2} = \frac{1}{2(1-\rho^2)}z_1^2(1-\rho^2)$$

so that the contents in brackets equals

$$\begin{aligned} -\frac{1}{2}\left[\frac{z_1^2 + z_2^2 - 2\rho z_1 z_2 - z_1^2(1-\rho^2)}{1-\rho^2}\right] &= -\frac{1}{2}\left[\frac{z_1^2(1-1+\rho^2) + z_2^2 - 2\rho z_1 z_2}{1-\rho^2}\right] \\ &= -\frac{1}{2(1-\rho^2)}\left[z_1^2\rho^2 + z_2^2 - 2\rho z_1 z_2\right] \\ &= -\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}\rho^2 + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right] \end{aligned}$$

Now, take out a common factor of $1/\sigma_2^2$:

$$\begin{aligned} &= -\frac{1}{2\sigma_2^2(1-\rho^2)}\left[\rho^2\frac{\sigma_2^2}{\sigma_1^2}(x_1-\mu_1)^2 + (x_2-\mu_2)^2 - 2\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)(x_2-\mu_2)\right] \\ &= -\frac{1}{2\sigma_2^2(1-\rho^2)}\left[x_2 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)\right]^2 && \text{(from (1))} \\ &= -\frac{1}{2}\left(\frac{x_2 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)}{\sigma_2\sqrt{1-\rho^2}}\right)^2 \end{aligned}$$

Therefore,

$$f_{X_2|X_1}(x_2|x_1) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho}^2} \exp\left\{-\frac{1}{2}\left(\frac{x_2 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)}{\sigma_2\sqrt{1-\rho^2}}\right)^2\right\},$$

as required. All that is left is to determine the distribution of $X_2|X_1$. That requires a bit of relabeling. In fact, if we define

$$\begin{aligned} \tilde{\mu}_2 &= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1) \\ \tilde{\sigma}_2^{2*} &= \sigma_2^2(1-\rho^2) \qquad \text{implying } \tilde{\sigma}_2 = \sigma_2\sqrt{1-\rho^2} \end{aligned}$$

then the density $f(x_2|x_1)$ can in fact be written as

$$f(x_2|x_1) = \frac{1}{\sqrt{2\pi\tilde{\sigma}_2^2}} \exp\left\{-\frac{(x_2 - \tilde{\mu}_2)^2}{2\tilde{\sigma}_2^2}\right\},$$

which matches the probability density function of a normally distributed random variable! So in fact we can say that the random variable $X_2|X_1$ is normally distributed with (conditional) mean $\tilde{\mu}_2$ and (conditional) variance $\tilde{\sigma}_2^2$. That is,

$$X_2|X_1 \sim N(\tilde{\mu}_2, \tilde{\sigma}_2^2).$$

Question 3: A definite problem

Note that for this question it is assumed that you know the definitions of (i) quadratic forms and (ii) eigenvalues and eigenvectors. It is also further assumed that you have knowledge of the property that positive (semi) definite matrices have eigenvalues which are strictly (weakly) positive, with analogous properties for negative (semi) definite matrices. If you don't know these, refer to the Economics Summer Math Intensive classes, or seek out the intensive instructor for further clarification.

(a) Recall that if a matrix A is p.s.d. then $x'Ax \geq 0$ for any $x \in \mathbb{R}^n$. Since A is symmetric, we can perform spectral decomposition on it so that

$$A = Q\Lambda Q'$$

where $Q^{-1} = Q'$. Observe that if we let $y = Q'x$ then we have that y is also a $n \times 1$ column vector which is not equal to zero. Therefore

$$\begin{aligned} x'Ax &= x'Q\Lambda Q'x \\ &= y'\Lambda y \\ &\geq 0 \end{aligned}$$

because A is positive semidefinite. However, this arrangement shows that Λ also satisfies the definition of positive semidefiniteness.

(b) The steps are similar but not completely identical to what we have above. First note that if A is negative definite then it is the case that $x'Ax > 0$ for any $x \neq 0$. Now, like before, we can decompose $A = Q\Lambda Q'$ and define $y = Q'x$ as before. But this time, we observe that Q is a nonsingular matrix. Coupled with the fact that $x \neq 0$, this implies that $y \neq 0$ *also.*

Now we have

$$\begin{aligned} x'Ax &= x'Q\Lambda Q'x \\ &= y'\Lambda y \\ &= \sum_{i=1}^{n} \lambda_i y_i^2 \\ &< 0 \end{aligned}$$

showing that Λ is also negative definite since all of the eigenvalues of A are negative (this uses the negative definite property of A) and at least one $y_i \neq 0$ (by virtue of $y \neq 0$). If you have trouble seeing how the summation above works, a simple two-variable case demonstrates this:

$$y'\Lambda y = \begin{bmatrix} y_1 & y_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \lambda_1 y_1^2 + \lambda_2 y_2^2.$$

(c) Since A is idempotent, then any eigenvector v of A satisfies the definition

$$\begin{aligned} Av = \lambda v \implies (AA)v &= \lambda v && \text{(Idempotence)} \\ \implies A\lambda v &= \lambda v \\ \implies \lambda A v &= \lambda v \\ \implies \lambda^2 v &= \lambda v \end{aligned}$$

which implies $\lambda^2 = \lambda$. This is true only when λ is either 1 or 0. Hence any eigenvectors of A has associated with it an eigenvalue that is either 1 or 0.[4]

(d) If A is full rank, then it is invertible. But A is also diagonalizable so that

$$\begin{aligned} A = Q\Lambda Q' \implies A^{-1} &= (Q\Lambda Q')^{-1} \text{ exists} \\ &= (Q')^{-1}\Lambda^{-1}Q^{-1} \\ &= Q\Lambda^{-1}Q' && \because Q' = Q^{-1}. \end{aligned}$$

But if the expression $Q\Lambda^{-1}Q'$ exists then Λ^{-1} must also exist. Therefore, Λ is also invertible.

[4] A different way to see this is to rearrange: $\lambda^2 - \lambda = 0 \implies \lambda(\lambda - 1) = 0$ from which $\lambda = 0, 1$ follows.