vrije Universiteit *amsterdam*

School of Business and Economics

Exam:	Multivariate Statistics
Code:	E_EOR2_MS
Examinator: Co-reader:	Andre Lucas Julia Schaumburg
Date:	May 20, 2019
Time:	12:00
Duration:	2 hours
Calculator allowed:	Yes
Graphical calculator allowed:	Yes
Number of questions:	4
Type of questions:	Open
Answer in:	English

Remarks: Take the number of points as a guidance on the maximum time to spend on that question.

Credit score:	See the credits for each separate question
Grades:	The grades will be made public latest on June 2, 2019.
Inspection:	Tuesday, June 4, 2019 at 10am. Room tba via CANVAS.
Number of pages:	6 (including front page)

Good luck!

Question 1. [40points]

You have two independent random variables, R and A, with $f_A(a) = \frac{1}{\pi}$for $-\frac{1}{2}\pi \leq a \leq \frac{1}{2}\pi$, and zero else, and $f_R(r) = \exp(-a)$ for $a > 0$ and zero else. We therefore have that A is a point on the half-circle, and R is its ray.

A) Define $Y_1 = R \cdot \cos(A)$ and $Y_2 = R \cdot \sin(A)$. Derive the multivariate cdf of $Y = (Y_1, Y_2)^\top$. [20pt] [Hint: it is convenient for deriving the expression of the inverse transformation to use the fact that R is the ray along the point of the circle (Pythagoras).]

Inverse transformation:

$\frac{Y_2}{Y_1} = \tan(A) \;\Leftrightarrow\; A = \tan^{-1}\left(\frac{Y_2}{Y_1}\right)$ [3pt]

$R = \frac{Y_2}{\sin\left(\tan^{-1}\left(\frac{Y_2}{Y_1}\right)\right)} = \frac{Y_1}{\cos\left(\tan^{-1}\left(\frac{Y_2}{Y_1}\right)\right)} = \sqrt{Y_1^2 + Y_2^2}$ [3pt]

Domain change: [4pt]

$$(Y_1, Y_2) \in \mathbb{R} \times (\mathbb{R}_+ \cup \{0\})$$

Jacobian of transformation (checked mapping is one on one): [5pt]

$$abs\left(\begin{vmatrix} \frac{-Y_2/Y_1^2}{1 + Y_2^2/Y_1^2} & \frac{1/Y_1}{1 + Y_2^2/Y_1^2} \\ \frac{Y_1}{\sqrt{Y_1^2 + Y_2^2}} & \frac{Y_2}{\sqrt{Y_1^2 + Y_2^2}} \end{vmatrix}\right) = \frac{1}{(Y_1^2 + Y_2^2)^{1/2}}$$

Pdf: [5pt]

$$\frac{1}{\pi} \cdot \exp\left(-\sqrt{Y_1^2 + Y_2^2}\right) \cdot \frac{1}{(Y_1^2 + Y_2^2)^{1/2}}$$

You have a normally distributed random vector $X = (X_1, X_2, X_3)^\top$, with

$$X_1|X_2 \sim N(X_2, 3), \qquad X_2|X_1 \sim N(1 + 0.25X_1, \star), \qquad X_3|\binom{X_1}{X_2} \sim N(X_2 - X_1, 1).$$

B) Provide the mean and covariance matrix of X. [20pt]
[Hint1: first solve for the distribution of (X_1, X_2).]
[Hint2: if you do not succeed in solving for the distribution of (X_1, X_2), then *(incorrectly) assume* $(X_1, X_2)^\top \sim N\left((1,2)^\top, \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}\right)$ and proceed to derive the distribution of X.]

Using the LIE, we get
$E[X_1] = E\big[E[X_1|X_2]\big] = E[X_2] = E[X_2]$, [1pt]
and similarly [1pt]

$$E[X_2] = 1 + 0.25E[X_1],$$

so [2pt] $E[X_1] = E[X_2] = 4/3$, and thus [2pt] $E[X_3] = E[X_2] - E[X_1] = 0$, so

$$\mu = \left(\frac{4}{3}, \frac{4}{3}, 0\right)^\top.$$

Using the expression for the conditional mean, we get [4pt]

$$\Sigma_{1,2}\Sigma_{2,2}^{-1} = 1, \qquad \Sigma_{2,1}\Sigma_{1,1}^{-1} = 0.25,$$

and thus $\Sigma_{2,2} = \Sigma_{1,2}$, $\Sigma_{1,1} = 4\Sigma_{1,2}$.
Using the expression for the conditional variance, we get [4pt]

$$\Sigma_{1,1} - \Sigma_{2,2} = 3 \Leftrightarrow 3\Sigma_{1,2} = 3 \Leftrightarrow \Sigma_{1,2} = 1$$

and thus $\Sigma_{1,1} = 4$ and $\Sigma_{2,2} = 1$, so [2pt]

$$\Sigma_{(1,2),(1,2)} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$$

And finally, from the expression for the conditional variance (using Th5.4) [2pt]

$$\Sigma_{3,(1,2)} = (-1,1)\Sigma_{(1,2),(1,2)} = (-3,0),$$

and [2pt]

$$\Sigma_{3,3} = 1 + (-1,1)\Sigma_{(1,2),(1,2)}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1 + 3 = 4.$$

So

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 4/3 \\ 4/3 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 & -3 \\ 1 & 1 & 0 \\ -3 & 0 & 4 \end{pmatrix}\right)$$

Question 2. [40pt]

A) Provide the algorithm to simulate from a 4 dimensional pdf that is characterized by $\chi^2(\nu_i)$-marginals for $i = 1, \ldots, 4$, and a Gaussian copula with correlation parameter $R \in \mathbb{R}^{4\times4}$. [10pt]

1. Draw $\tilde{X}$ from a multivariate normal with mean zero and covariance matrix R.
2. Transform each of the elements of $\tilde{X}_i$ for $i = 1, \ldots, 4$, into its PIT $U_i = \Phi(\tilde{X}_i)$, where $\Phi(\cdot)$ is the standard normal cdf.
3. Define $X_i = F_{X_i}^{-1}(U_i)$, where $F_{X_i}^{-1}(\cdot)$ is the inverse cdf of a $\chi^2(\nu_i)$ distribution.

Let $C^G(u_1, u_2, u_3; R)$ be the trivariate Gaussian copula with parameter R, and $C^{St}(u_1, u_2, u_3; R, \nu)$ the Student's t copula with parameters R and ν.
You construct a new copula as a mixture of these previous two, with mixture weights $w = 0.2$ and $1 - w = 0.8$, respectively.

B) Describe the algorithm to simulate from this mixture of the two copulas. [5pt]
[Hint: just concentrate in how to draw from a mixture distribution. Skip how to draw from a copula, which was question 2A above]

1. Draw a Bernoulli with success probability $w = 0.2$.
2. If success, draw from the Gaussian copula, otherwise, from the Student's t copula.

C) Show whether or not the mixture of the two copulas

$$C(u_1, u_2, u_3; R, \nu, w) = 0.2 \cdot C^G(u_1, u_2, u_3; R) + 0.8 \cdot C^{St}(u_1, u_2, u_3; R, \nu)$$

is itself a copula or not. [15pt]

1. It is grounded, as

$$C(u_1, u_2, 0; R, \nu, w) = 0.2 \cdot C^G(u_1, u_2, 0; R) + 0.8 \cdot C^{St}(u_1, u_2, 0; R, \nu) =$$

$$0.2 * 0 + 0.8 * 0 = 0$$
as the Gaussian and Student's t copula are grounded. Same holds for the other u_i.
[5pt]

2. The copula density is non-negative, as
$$c(u_1, u_2, 0; R, \nu, w) = 0.2 \cdot c^G(u_1, u_2, 0; R) + 0.8 \cdot c^{St}(u_1, u_2, 0; R, \nu) \geq 0,$$
as the Gaussian and Student's t copula densities are non-negative everywhere on the unit square.
[5pt]

3. It has uniform marginals, as
$$C(u_1, 1, 1; R, \nu, w) = 0.2 \cdot C^G(u_1, 1, 1; R) + 0.8 \cdot C^{St}(u_1, 1, 1; R, \nu) = 0.2u_1 + 0.8u_1 = u_1,$$
as the Gaussian and Student's t copula have uniform marginals. Same holds for the other u_i.
[5pt]

So yes, it is a copula.

Assume the marginal distributions
$$f_{X_1}(x_1) = 2 \cdot \exp(-2x_1), \qquad F_{X_1}(x_1) = 1 - \exp(-2x_1),$$
$$f_{X_2}(x_2) = 4 \cdot \exp(-4x_2), \qquad F_{X_2}(x_2) = 1 - \exp(-4x_2),$$
$$f_{X_3}(x_3) = 3 \cdot \exp(-3x_3), \qquad F_{X_3}(x_3) = 1 - \exp(-3x_3),$$
and assume that $X = (X_1, X_2, X_3)^\top$ has the mixture copula dependence structure from question (C).

D) Given a realization $X = x = (2,2,2)^\top$, how can you compute the probability that this observation comes from the *second* (i.e., Student's t copula) mixture component (given the outcome)? Provide the equation. [10pt]

$u_1 = 1 - e^{-4}$, $u_2 = 1 - e^{-8}$, $u_3 = 1 - e^{-6}$, [5pt]
Then the probability is given by [5pt]
$$\frac{0.8 \cdot C^{St}(u_1, u_2, u_3; R, \nu)}{0.2 \cdot C^G(u_1, u_2, u_3; R) + 0.8 \cdot C^{St}(u_1, u_2, u_3; R, \nu)}$$

Question 3. [15pt]

On a sample of size $n = 101$ you estimated
$$\bar{x} = (1 \;\; 4.9 \;\; 5.5 \;\; 0.7)^\top,$$
$$\mathcal{S} = \begin{pmatrix} 7 & 4 & -2 & -2 \\ 4 & 6 & -1 & 0 \\ -2 & -1 & c & 1 \\ -2 & 0 & 1 & 5 \end{pmatrix},$$
for some $c > 0$, where $\bar{x}$ and $\mathcal{S}$ denote the sample mean and covariance, respectively.
You formulate the joint hypothesis that the means of X_1 and X_4 are equal, *and* that the means of X_2 and X_3 are equal.

χ^2 critical values for ν degrees of freedom	5%
$\nu = 1$	3.84
$\nu = 2$	5.99

$\nu = 3$	7.81
$\nu = 4$	9.49

A) At a 5% significance level, how large must c be for you to *not reject* your hypothesis using a χ^2 test? [15pt]

$$\mathcal{A} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{pmatrix}$$
$$H0: \mathcal{A}\mu = 0$$

[4pt]

$\mathcal{A}\mu = \begin{pmatrix} 0.3 \\ -0.6 \end{pmatrix}$ [2pt]

$\mathcal{A}S\mathcal{A}^\top = \begin{pmatrix} 9 & 4 & -3 & -7 \\ 6 & 7 & -1-c & -1 \end{pmatrix}\mathcal{A}^\top = \begin{pmatrix} 16 & 7 \\ 7 & 8+c \end{pmatrix}$ [3pt]

$$T = \frac{100}{79+16c}(0.3 \quad -0.6)\begin{pmatrix} 8+c & -7 \\ -7 & 16 \end{pmatrix}\begin{pmatrix} 0.3 \\ -0.6 \end{pmatrix} = \frac{900+9c}{79+16c} \text{ [4pt]} < 5.99 \text{ [1pt]}$$
$$900 + 9c < 5.99(79 + 16c)$$

$c > 4.91$. [1pt]

Question 4. [25pt]

A) In a PCA, explain why it is important to normalize the variables in terms of their means and standard deviations. [5pt]

Otherwise, the PCA factors will just pick up the variable with the largest variance first, regardless of any hidden components. (or similarly with the largest mean)

B) Describe the main steps of one of the iterative (non-maximum-likelihood) techniques to estimate an orthogonal factor model. [10pt]

Choose one of the two from the book.

You estimated an orthogonal factor model and obtained the factor loading matrix

$$L = \begin{pmatrix} 2 & 3 \\ -1 & 1 \\ 1 & 4 \end{pmatrix}.$$

Your colleague says that you can multiply the factor loadings from the right by the matrix $\mathcal{A}$ with

$$\mathcal{A} = \begin{pmatrix} 0.2 & -0.6 \\ 0.2 & 0.4 \end{pmatrix}$$

and obtain

$$\tilde{L} = L \cdot \mathcal{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix},$$

and that this model yields an orthogonal factor model that is indistinguishable from your model, and that therefore you can interpret the first factor as the average of the 1st and 3rd variable, and the second factor as the average of the 2nd and 3rd factor.

C) Comment whether your colleague is right or wrong. [10pt]

Wrong as $\mathcal{A}$ is not an orthogonal matrix, and therefore the new model is *not* an orthogonal factor model.
As an aside, the factor is not an average of the two variables, as this discards the idiosyncratic error term that applies to each variable.