STAT 30100 : Mathematical Statistics-1

Winter 2020

•

Homework 2

•

Solutions by

Jinhong Du

12243476

STAT 30100, Homework 2

1. (a) Suppose $\{X_1, \dots, X_k\}$, $k \geq 2$, are identically distributed with finite first and second moments and common pairwise correlation $r = \mathrm{Corr}(X_i, X_j)$ for $i \neq j$. Prove $r \geq -\frac{1}{k-1}$.

Proof. Let $\sigma^2 = \mathrm{Var}(X_i)$.
If $\sigma^2 = 0$, then $X_1 = \cdots = X_k = c$ almost surely for some $c \in \mathbb{R}$. In this case, the correlation is undefined.
If $\sigma^2 > 0$, since $r = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, Y_i)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(Y_i)}}$, we have $\mathrm{Cov}(X_i, Y_i) = r\sigma^2$. Since

$$
\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) &= \sum_{i=1}^{k} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, Y_i) \\
&= k\sigma^2 + 2\frac{(1+k-1)(k-1)}{2}\sigma^2 \\
&= \sigma^2[k + k(k-1)r] \geq 0,
\end{aligned}
$$

we have $r \geq -\frac{1}{k-1}$. □

(b) Under the setting of (a), prove that the correlation attains its lower bound (i.e. $r = -\frac{1}{k-1}$) if and only if $\sum_{i=1}^{k} X_i =$ constant almost surely.

Proof. $r = -\frac{1}{k-1}$ if and only if $\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) = 0$, if and only if $\sum_{i=1}^{k} X_i =$ constant almost surely. □

2. Prove the following result, known as the Hoeffding identity. Let $F(x, y)$ be the joint c.d.f. of (X, Y). Suppose $\mathrm{Cov}(X, Y)$ exists, then

$$
\mathrm{Cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [F(x, y) - F(x, +\infty)F(+\infty, y)]\mathrm{d}x\mathrm{d}y.
$$

Proof. Let (X', Y') be an independent copy of (X, Y). Then

$$
\mathbb{E}[(X - X')(Y - Y')] = \mathbb{E}(XY) - \mathbb{E}(X')\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y') + \mathbb{E}(X'Y') = 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)].
$$

Therefore,

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\
&= \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y \\
&= \mathbb{E}[(X - X')(Y - Y')] \\
&= \mathbb{E}\left[\int_{\mathbb{R}} (\mathbb{1}_{u \leq X} - \mathbb{1}_{u \leq X'})\mathrm{d}u \cdot \int_{\mathbb{R}} (\mathbb{1}_{v \leq Y} - \mathbb{1}_{v \leq Y'})\mathrm{d}v\right]
\end{aligned}
$$

Solution (cont.)

$$= \mathbb{E}\left[\iint_{\mathbb{R}^2} (\mathbb{1}_{u\le X} - \mathbb{1}_{u\le X'})(\mathbb{1}_{v\le Y} - \mathbb{1}_{v\le Y'})\mathrm{d}u\mathrm{d}v\right]$$
$$= \iint_{\mathbb{R}^2} \mathbb{E}[(\mathbb{1}_{u\le X} - \mathbb{1}_{u\le X'})(\mathbb{1}_{v\le Y} - \mathbb{1}_{v\le Y'})]\mathrm{d}u\mathrm{d}v,$$

where the last equality comes from Fubini Theorem. Since

$$\begin{aligned}
&\mathbb{E}[(\mathbb{1}_{u\le X} - \mathbb{1}_{u\le X'})(\mathbb{1}_{v\le Y} - \mathbb{1}_{v\le Y'})]\\
=&\mathbb{E}[\mathbb{1}_{u\le X}\mathbb{1}_{v\le Y} - \mathbb{1}_{u\le X}\mathbb{1}_{v\le Y'} - \mathbb{1}_{u\le X'}\mathbb{1}_{v\le Y} + \mathbb{1}_{u\le X'}\mathbb{1}_{v\le Y'}]\\
=&\mathbb{P}(X\ge u, Y\ge v) - [1-F(u,+\infty)][1-F(+\infty,v)]\\
=&F(u,+\infty)+F(+\infty,v)+[1-\mathbb{P}(X\ge u, Y\ge v)] - F(u,+\infty)F(+\infty,v)\\
=&F(u,+\infty)+F(+\infty,v)-\mathbb{P}(\{X<u \text{ or } Y<v\}) - F(u,+\infty)F(+\infty,v)\\
=&F(u,v)-F(u,+\infty)F(+\infty,v),
\end{aligned}$$

we have $\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F(u,v)-F(u,+\infty)F(+\infty,v)]\mathrm{d}u\mathrm{d}v$. □

3. Use the Hoeffding identity to prove that for any functions $g(x)$ and $h(y)$ such that $\mathrm{Cov}(g(X),h(Y))$ exists we have
$$\mathrm{Cov}(g(X),h(Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathrm{Cov}(\mathbb{1}_{(s,+\infty)}(g(X)),\mathbb{1}_{(t,+\infty)}(h(X)))\mathrm{d}t\mathrm{d}s,$$
where $\mathbb{1}_A(\cdot)$ is the indicator function for set A.

Proof. Let $F(x,y)$ be the joint distribution function of $h(X)$ and $h(Y)$. Then from the Hoeffding identity, we have

$$\mathrm{Cov}(g(X),h(Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F(s,t)-F(s,+\infty)F(+\infty,t)]\mathrm{d}s\mathrm{d}t.$$

Since

$$\begin{aligned}
&\mathrm{Cov}(\mathbb{1}_{(s,+\infty)}(g(X)),\mathbb{1}_{(t,+\infty)}(h(X)))\\
=&\mathbb{E}[\mathbb{1}_{(s,+\infty)}(g(X))\mathbb{1}_{(t,+\infty)}(h(X))] - \mathbb{E}[\mathbb{1}_{(s,+\infty)}(g(X))]\mathbb{E}[\mathbb{1}_{(t,+\infty)}(h(X))]\\
=&\mathbb{P}(g(X)\ge s, h(Y)\ge t) - [1-F(s,+\infty)][1-F(+\infty,t)]\\
=&\mathbb{P}(g(X)\ge s, h(Y)\ge t) + F(s,+\infty)+F(+\infty,t)-1-F(s,+\infty)F(+\infty,t)\\
=&F(s,t)-F(s,+\infty)F(+\infty,t),
\end{aligned}$$

we have
$$\mathrm{Cov}(g(X),h(Y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\mathrm{Cov}(\mathbb{1}_{(s,+\infty)}(g(X)),\mathbb{1}_{(t,+\infty)}(h(X)))\mathrm{d}t\mathrm{d}s.$$

□

4. Use the identity in the previous problem to prove that $\mathrm{Cov}(g(X), h(Y)) \le 0$ for all nondecreasing functions g and h if and only if $\mathrm{Cov}(g(X), h(Y)) \le 0$ for all binary (i.e. functions that take only two values) nondecreasing functions g and h.

Proof. $\Longleftarrow$ Since for all binary nondecreasing functions g' and h', $\mathrm{Cov}(g'(X), h'(Y)) \le 0$, we have that for any nondecreasing function g and h, any $s, t \in \mathbb{R}$,

$$\mathrm{Cov}(\mathbb{1}_{(s,+\infty)}(g(X)), \mathbb{1}_{(t,+\infty)}(h(X))) \le 0.$$

Therefore, $\mathrm{Cov}(g(X), h(Y)) \le 0$.
$\Longrightarrow$ Obviously since $\mathrm{Cov}(g(X), h(Y)) \le 0$ for all nondecreasing functions g and h, implies $\mathrm{Cov}(g(X), h(Y)) \le 0$ for all binary nondecreasing functions g and h. $\square$

5. Does there always exist (i.e., for all choices of $a \ge 0$, $b > 0$, and $c > 0$) a probability distribution for the random vector (X, Y) with support $(0, \infty) \times (0, \infty)$ such that for $y > 0$, the conditional density of X given $Y = y$ is

$$f(x|y) = (ay + b) \exp[-(ay + b)x], \qquad x > 0,$$

and for $x > 0$, the conditional density of Y given $X = x$ is

$$g(y|x) = (ax + c) \exp[-(ax + c)y], \qquad y > 0,$$

where $a \ge 0$, $b > 0$, and $c > 0$ are fixed parameters? Prove your answer by checking the conditions given in class.

(1) Functional compatibility:

$$\frac{f(x|y)}{g(y|x)} = \frac{1}{(cx + d)e^{bx}} \cdot (ay + b)e^{cy}$$

(2) For all $y > 0$,

$$\int_0^{+\infty} \frac{f(x|y)}{g(y|x)} \mathrm{d}x = (ay + b)e^{cy} \int_0^{+\infty} \frac{1}{(cx + d)e^{bx}} \mathrm{d}x < \infty$$
$$\int_0^{+\infty} \frac{g(y|x)}{f(x|y)} \mathrm{d}y = (cx + d)e^{bx} \int_0^{+\infty} \frac{1}{(ay + b)e^{cy}} \mathrm{d}y < \infty$$

Therefore, there exists such a probability distribution.

6. (Problem 4.48 in Casella and Berger) Gelman and Meng (1991) give an example of a bivariate family of distributions that are not bivariate normal but have normal conditionais. Define the joint pdf of (X, Y) as

$$f(x,y) \propto \exp\left\{-\frac{1}{2}[Ax^2y^2 + x^2 + y^2 - 2Bxy - 2Cx - 2Dy]\right\},$$

where A, B, C, D are constants.

(a) Show that the distribution of $X|Y = y$ is normal with mean $\frac{By+C}{Ay^2+1}$ and variance $\frac{1}{Ay^2+1}$. Derive a corresponding result for the distribution of $Y|X = x$.

Proof. Let $c_0 = \iint_{\mathbb{R}^2} \exp\left\{-\frac{1}{2}[Ax^2y^2 + x^2 + y^2 - 2Bxy - 2Cx - 2Dy]\right\} \mathrm{d}x\mathrm{d}y$, then $f(x,y) = c_0^{-1} \exp\left\{-\frac{1}{2}[Ax^2y^2 + x^2 + y^2 - 2Bxy - 2Cx - 2Dy]\right\}$.

$$\begin{aligned} f_{X|Y=y}(x) &= c_0^{-1} e^{-\frac{1}{2}\left[y^2-2Dy\right]} e^{-\frac{1}{2}\left[(Ay^2+1)x^2-2(By+C)x\right]} \\ &= c_1^{-1} e^{-\frac{1}{2}\left[(Ay^2+1)\left(x-\frac{By+C}{Ay^2+1}\right)^2\right]} \end{aligned}$$

where $c_1 = c_0 e^{\frac{1}{2}\left[y^2-2Dy-\frac{(By+C)^2}{Ay^2+1}\right]}$ is a normalizing factor. Therefore, $X|Y = y$ is normal with mean $\frac{By+C}{Ay^2+1}$ and variance $\frac{1}{Ay^2+1}$.

$$\begin{aligned} f_{Y|X=x}(y) &= c_0^{-1} e^{-\frac{1}{2}\left[x^2-2Cx\right]} e^{-\frac{1}{2}\left[(Ax^2+1)y^2-2(Bx+D)y\right]} \\ &= c_2^{-1} e^{-\frac{1}{2}\left[(Ax^2+1)\left(y-\frac{Bx+D}{Ax^2+1}\right)^2\right]} \end{aligned}$$

where $c_2 = c_0 e^{\frac{1}{2}\left[x^2-2Cx-\frac{(Bx+D)^2}{Ax^2+1}\right]}$ is a normalizing factor. Therefore, $Y|X = x$ is normal with mean $\frac{Bx+D}{Ax^2+1}$ and variance $\frac{1}{Ax^2+1}$. □

(b) A most interesting configuration is $A = 1$, $B = 0$, $C = D = 8$. Show that this joint distribution is bimodal.

Proof. When $A = 1$, $B = 0$, $C = D = 8$,

$$f(x,y) \propto \exp\left\{-\frac{1}{2}[x^2y^2 + x^2 + y^2 - 16x - 16y]\right\}.$$

Let $h(x,y) = x^2y^2 + x^2 + y^2 - 16x - 16y$. Since

$$\frac{\partial h(x,y)}{\partial x} = 2x(y^2+1) - 16 = 0, \tag{1}$$

$$\frac{\partial h(x,y)}{\partial y} = 2y(x^2+1) - 16 = 0, \tag{2}$$

by (1)-(2) we have

$$(xy-1)(y-x) = 0. \tag{3}$$

From (1) we also have that the solutions $x, y > 0$.
If $xy = 1$, we have $y^2 + 1 = 8y$, i.e., $y = 4 \pm \sqrt{15}$ and so $x = 4 \mp \sqrt{15}$. $h(4 \pm \sqrt{15}, 4 \mp \sqrt{15}) = -65$.
If $x = y$, we have $x^3 + x = 8$. Since $\frac{\mathrm{d}(x^3+x)}{\mathrm{d}x} = 3x^2 + 1 > 0$ when $x > 0$ and $2^3 + 2 = 10 > 8$, the solutions of $x^3 + x = 8$ must be less than 2. $h(2,2) = -64 \geq -65$.
So, h has two minimums. Therefore, $f(x,y)$ has two maximums, i.e., this joint distribution is bimodal. □

7. Suppose $X \sim N_3(0, M)$, where M is the matrix

$$\begin{pmatrix} \frac{7}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{5}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{7}{6} \end{pmatrix}$$

Let $Z = (Z_1, Z_2, Z_3)$ satisfy $Z \sim N_3(0, I)$. Find a vector of linear combinations of Z_1, Z_2, and Z_3 that has the same distribution as X.

Proof. Since M is a symmetric matrix, it has an unitary eigen-decomposition, $M = U\Sigma U^\top$ where

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 2 \end{pmatrix}.$$

Let $T = U\Sigma^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}$, we have $TT^\top = M$. Therefore, $TZ \sim N_3(0, TT^\top) = N_3(0, M)$, i.e., $TZ \overset{D}{=\!=} X$. Then the vector is given by

$$TZ = \begin{pmatrix} \frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{3}} Z_2 + \frac{1}{\sqrt{3}} Z_3 \\ \frac{1}{\sqrt{3}} Z_2 - \frac{2}{\sqrt{3}} Z_3 \\ -\frac{1}{\sqrt{2}} Z_1 + \frac{1}{\sqrt{3}} Z_2 + \frac{1}{\sqrt{3}} Z_3 \end{pmatrix}.$$

$\square$

8. Let $X \sim N_n(\mu, \Sigma)$, $r(\Sigma) = n$. Find the conditional distribution of X_1 given $X_2, X_3, \ldots, X_n$ directly from the density (please do not regurgitate the alternative approach I gave in class). Express your answer in terms of $\mu_1, \ldots, \mu_n$ and Σ_{11} (a scalar), Σ_{10} (of dimension $1 \times (n-1)$), Σ_{01} (of dimension $(n-1) \times 1$), Σ_{00} (an $(n-1) \times (n-1)$ matrix), where $\mu = (\mu_1, \ldots, \mu_n)^\top$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{10} \\ \Sigma_{01} & \Sigma_{00} \end{pmatrix}$.

[Hint: If you first establish that the density is normal by finding the form of the density, you can then try to determine which normal distribution it is (i.e. the mean and variance) by noting that a normal density is at a maximum at the mean and the variance is related in a simple way to the coefficient of x^2 in the density. Formula (38) on the matrix review sheet may be useful.]

Proof. From formula (38) on the matrix review, we have

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + FE^{-1}F^\top & -FE^{-1} \\ -E^{-1}F^\top & E^{-1} \end{pmatrix}$$

where $E = \Sigma_{00} - \Sigma_{10}^\top \Sigma_{11} \Sigma_{10} \in \mathbb{R}^{(n-1)\times(n-1)}$ and $F = \Sigma_{11}^{-1}\Sigma_{10} \in \mathbb{R}^{1\times(n-1)}$.
The joint density is given by

$$\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

Let $z_1 = x_1 - \mu_1$ and $z_0 = (x_2 - \mu_2, \ldots, x_n - \mu_n)^\top$, we have

$$\begin{aligned}
&(x-\mu)^\top \Sigma^{-1}(x-\mu) \\
=&z_1^2(\Sigma_{11}^{-1} + FE^{-1}F^\top) - 2z_1FE^{-1}z_0 + z_0^\top E^{-1}z_0 \\
=&(\Sigma_{11}^{-1} + FE^{-1}F^\top)\left(z_1 - \frac{FE^{-1}z_0}{\Sigma_{11}^{-1} + FE^{-1}F^\top}\right)^2 - \frac{(FE^{-1}z_0)^2}{\Sigma_{11}^{-1} + FE^{-1}F^\top} + z_0^\top E^{-1}z_0 \\
=&(\Sigma_{11}^{-1} + FE^{-1}F^\top)\left(x_1 - \mu_1 - \frac{FE^{-1}z_0}{\Sigma_{11}^{-1} + FE^{-1}F^\top}\right)^2 - \frac{(FE^{-1}z_0)^2}{\Sigma_{11}^{-1} + FE^{-1}F^\top} + z_0^\top E^{-1}z_0
\end{aligned}$$

i.e., $f_{X_1|X_2=x_2,\ldots,X_n=x_n}(x)$ the conditional density of X_1 given $X_2, \ldots, X_n$ is proportional to

$$\exp\left\{-\frac{1}{2}(\Sigma_{11}^{-1} + FE^{-1}F^\top)\left(x_1 - \mu_1 - \frac{FE^{-1}z_0}{\Sigma_{11}^{-1} + FE^{-1}F^\top}\right)^2\right\}.$$

Therefore, $f_{X_1|X_2=x_2,\ldots,X_n=x_n}(x)$ is normal with mean $\mu_1 + \frac{FE^{-1}z_0}{\Sigma_{11}^{-1}+FE^{-1}F^\top}$ and variance $(\Sigma_{11}^{-1} + FE^{-1}F^\top)^{-1}$, where $E = \Sigma_{00} - \Sigma_{10}^\top\Sigma_{11}\Sigma_{10}$, $F = \Sigma_{11}^{-1}\Sigma_{10}$ and $z_0 = (x_2 - \mu_2, \ldots, x_n - \mu_n)^\top$. $\square$