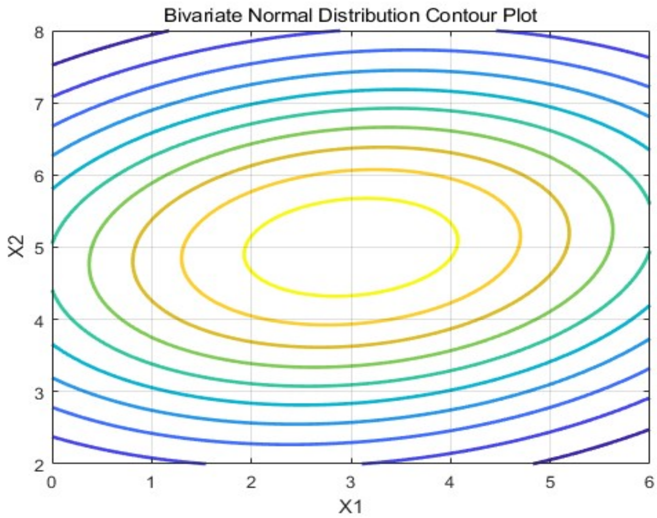
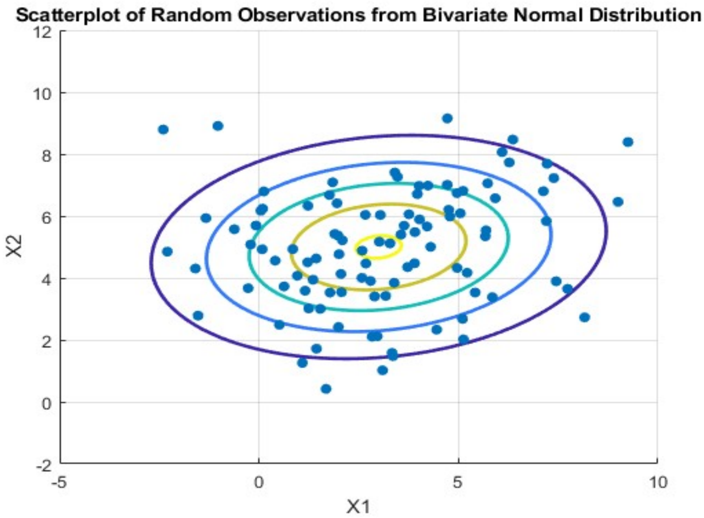
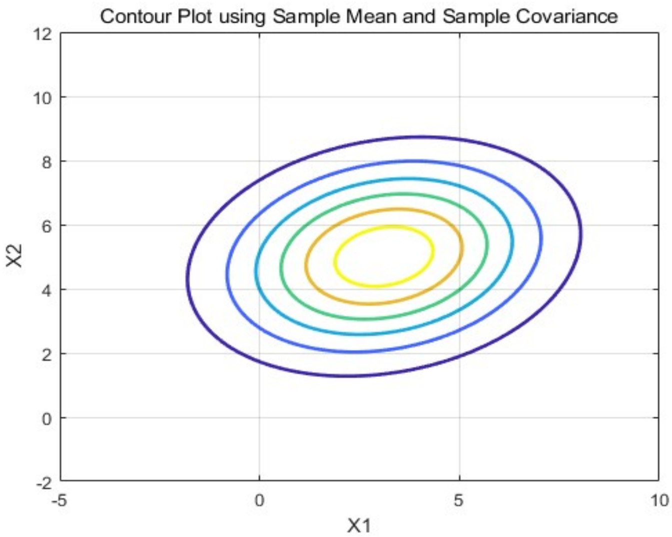




Sample Mean:
3.1142 5.0032

Population Mean:
3 5

Sample Covariance:
6.3302 0.9063
0.9063 3.6319



Q2. $(X_1, X_2)' \sim N_2\left(\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}4 & 0\\0 & 5\end{bmatrix}\right)$

Q3. (a) $\sigma_{12}^2 = \sigma_{21}^2 = 0 \Rightarrow$ independent X_1 & X_2

(b) $\sigma_{13}^2 = \sigma_{31}^2 = -1 \Rightarrow X_1$ & X_3 are dependent

(c) $\vec{X}' = \begin{pmatrix}X_1\\X_3\\ \hline X_2\end{pmatrix}$, $\Sigma' = \left[\begin{array}{cc|c}4 & -1 & 0\\-1 & 2 & 0\\ \hline 0 & 0 & 5\end{array}\right]$ $\therefore$ (X_1, X_3) & X_2 are independent (covariate $=0$)

(d) Set $A = \begin{pmatrix}1 & 0 & 0\\1 & 3 & -2\end{pmatrix}$ $\quad \Sigma' = A\Sigma A^T$, the covariance of $\begin{pmatrix}X_1\\X_1+3X_2-2X_3\end{pmatrix}$ $\Sigma' = \begin{pmatrix}4 & 6\\6 & 61\end{pmatrix} \Rightarrow$ dependent

Q4. $E(X_K) = 0 \quad E(\rho X_K + \varepsilon_{K+1}^N) = \rho E(X_K) + E(\varepsilon_N) = 0.$

$Var(X_K) = \frac{1}{1-\rho^2} \quad Var(\rho X_K + \varepsilon_{K+1}^N) = \rho^2 \cdot Var(X_K) + Var(\varepsilon_N) = \frac{\rho^2}{1-\rho^2} + 1 = \frac{1}{1-\rho^2}$

(a) $\because E(X_1) = 0$, $Var(X_1) = \frac{1}{1-\rho^2}$ $\therefore E(X_K) = 0$, $Var(X_K) = \frac{1}{1-\rho^2}$ $\forall K \in \mathbb{Z}$.

$Cov(X_K, X_{K+1}) = E(X_K - E(X_K))(X_{K+1} - E(X_{K+1})) = E(X_K \cdot X_{K+1}) = E(\rho X_K^2 + X_K \cdot \varepsilon_{K+1})$

$Cov(X_K, X_{K+1}) = \sigma_{K,K+1}^2 = \frac{\rho}{1-\rho^2}$ similar, $\sigma_{K,K+a}^2 = \frac{\rho^a}{1-\rho^2}$ $\quad = \rho E(X_K^2) = \rho \cdot (\sigma_K^2 + 0^2)$

$\therefore X \sim N(0_n, \Sigma_\rho)$ where $0_n = \begin{pmatrix}0\\ \vdots\\0\end{pmatrix} \in \mathbb{R}^n$ $(\Sigma_\rho)_{ij} = \frac{\rho^{|i-j|}}{1-\rho^2}$

(b) $X_2 = \rho X_1 + \varepsilon_2 \Rightarrow X_1 = \frac{X_2 - \varepsilon_2}{\rho}$ $\quad X_3 = \rho X_2 + \varepsilon_3$. $\varepsilon_2 \perp\!\!\!\perp \varepsilon_3$

$\therefore Cov(X_1, X_3)|X_2 = x_2 = E\left[(X_1 - E(X_1))(X_3 - E(X_3)) \mid X_2 = x_2\right] = E\left(\frac{X_2-\varepsilon_2}{\rho} - \frac{X_2}{\rho}\right)(\rho X_2 + \varepsilon_3 - \rho X_2)$

$= E\left(\frac{\varepsilon_2}{\rho} \cdot \varepsilon_3\right) = \frac{1}{\rho} E(\varepsilon_2 \cdot \varepsilon_3)$. $\varepsilon_2 . \varepsilon_3 \overset{i.i.d}{\sim} N(0,1)$, $LHS = 0$

Conditional independent

Q5. $X = (\vec{X_1}, \vec{X_2} \cdots \vec{X_n})$ $\quad \overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} \vec{X_i} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix}X_{1i}\\X_{2i}\\ \vdots\\X_{pi}\end{pmatrix}$ $\quad D = diag(s_i^2) = \frac{1}{n-1} diag(SST_1, SST_2, \cdots, SST_n)$

(a) $T = \frac{n}{p}\overline{X}_n^T D^{-1} \overline{X}_n = \frac{n}{p}\begin{pmatrix}\overline{X_1}\\\overline{X_2}\\ \vdots\\ \overline{X_n}\end{pmatrix}^T \begin{pmatrix}\frac{n-1}{SST_1} & & \\ & \ddots & \\ & & \frac{n-1}{SST_n}\end{pmatrix}\begin{pmatrix}\overline{X_1}\\ \vdots\\ \overline{X_n}\end{pmatrix} = \frac{n}{p}\sum_{i=1}^{p}\left(\frac{(n-1)\,\overline{X_i}^2}{\sum_{j=1}^{n}(X_{ji} - \overline{X_i})^2}\right) = \frac{1}{p}\sum_{i=1}^{p} Y_i$

~~$Y_i \sim t_{n-1}$~~ $Y_i \sim F_{1,n-1}$, T has no regonizable form. (SST_i) s.t. $Y_i = \frac{n\overline{X_i}^2}{S_i^2}$

(b) $\det(D) = \prod_{i=1}^{n} \cancel{(n-1)} s_i^2 \neq 0 . \Rightarrow D^{-1}$ exists, $T = \frac{n}{p}\overline{X}' D^{-1}\overline{X}$ exists and determined.

Well defined.

(c) $Y_i = \frac{n(\overline{X}_i^2)}{\frac{\Sigma(X_{ji} - \overline{X_i})^2}{n-1}} \overset{iid}{\sim} F_{1,n-1}$ $\quad T = \frac{1}{p}\sum_{i=1}^{p} Y_i = \overline{Y}_p \xrightarrow{CLT} \sqrt{p}(\overline{Y}_p - E(Y_i)) \to N(0, \sigma_{Y_i}^2)$

$\overline{Y}_p - E(Y_i) \xrightarrow{P} N\left(0, \frac{\sigma_{Y_i}^2}{p}\right)$ when p is large enough.

Statistics 2221
Advanced Applied Multivariate Analysis
Spring 2024
Homework 2

Assigned: Saturday, February 3, 2024
Due date: In canvas, by end of day, Tuesday, February 13, 2024

1. Consider a bivariate normal distribution for $\mathbf{X} = (X_1, X_2)'$, with $\mathbb{E}(X_1) = 3$, $\mathbb{E}(X_2) = 5$, $\mathrm{Var}(X_1) = 10$, $\mathrm{Var}(X_2) = 4$, $\mathrm{Corr}(X_1, X_2) = 0.9$.
   (a) Sketch a contour of the density.
   (b) Generate $n = 100$ random observations from this distribution, and draw a scatterplot of the observations.
   (c) Report the sample mean, and compare it with the population mean.
   (d) Report the sample covariance matrix $\mathbf{S}$, and sketch the equidistance set given by the sample Mahalanobis distance, i.e.,
   $$E_c = \{\mathbf{y} \in \mathbb{R}^2 : (\mathbf{y} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{y} - \bar{\mathbf{x}}) = c^2\}$$
   for a $c > 0$.

2. Let $\mathbf{X} = (X_1, X_2, X_3)'$ be distributed as $N_3(\mu, \Sigma)$, where $\mu = (1, -1, 2)'$ and
   $$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$
   What is the distribution of a random vector $(X_1, X_2)'$?

3. Refer to the previous problem. Which of the following random variables are independent? Explain.
   (a) X_1 and X_2
   (b) X_1 and X_3
   (c) (X_1, X_3) and X_2
   (d) X_1 and $X_1 + 3X_2 - 2X_3$

4. Consider a random vector $\mathbf{X} = (X_1, \ldots, X_N)'$ for some natural number N. The joint distribution of $X_1, \ldots, X_N$ is given by the following relation:
   $$\begin{aligned} X_1 &= \epsilon_1, \epsilon_1 \sim N(0, \frac{1}{1-\rho^2}) \\ X_2 &= \rho X_1 + \epsilon_2, \epsilon_2 \sim N(0,1) \\ &\vdots \\ X_N &= \rho X_{N-1} + \epsilon_N, \epsilon_N \sim N(0,1), \end{aligned}$$
   where $0 < \rho < 1$ is fixed, and $(\epsilon_1, \ldots, \epsilon_N)$ are independent with each other.

(a) What is the distribution of $\mathbf{X}$? Specify its mean and variance-covariance matrix. [Show this for the $N = 3$ case.]

(b) Show that X_1 and X_3 are conditionally independent given $X_2 = x_2$.

Note: When $X_1, \ldots, X_N$ are viewed as a sequence of random variables (ordered by time), the above relation leads to an autoregressive model of time series.

The following are extra credit questions for non-statistics students. Graduate students in statistics are required to do the problems.

5. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from $N_p(\mathbf{0}, \Lambda)$, where Λ is a $p \times p$ diagonal matrix of $\sigma_1^2, \ldots, \sigma_p^2$. Let $D = \text{diag}(s_1^2, \ldots, s_p^2)$, for s_i^2 being the ith diagonal element of the sample variance–covariance matrix. Define a modified Hotelling's statistic $T = \frac{n}{p}\bar{X}'D^{-1}\bar{X}$.

   (a) Does T have a recognizable distribution?

   (b) Argue that T is well-defined even when $n < p$. (Hint: consider the invertibility of D)

   (c) Show that the distribution of T can be approximated by a normal distribution, if p is large enough. [Hint: Use central limit theorem for large p.]