Multivariate Normal Distribution Simulation

2020111538 Oh SukJu

Contents

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \right)$$

Let $Y \perp X_1$

then, $Y = C \cdot \ X_1 + X_2$

since $\text{Cov}(Y|X_1) = 0$,

$C \cdot \ \text{Var}(X_1) + \text{Cov}(X_1, X_2) = 0$

$C = -\sigma_{12} \cdot \sigma_{11}^{-1}$

So, $Y = X_2 - \sigma_{12} \cdot \ \sigma_{11}^{-1} X_1$

Mean and Variance of $X_2|X_1$

$$\begin{aligned} E[X_2|X_1] &= E[Y + \sigma_{12}\sigma_{11}^{-1}X_1|X_1] \\ &= E[Y|X_1] + \sigma_{12} \cdot \sigma_{11}^{-1}x_1 \\ &(Y \perp X_1) \\ &= \mu_2 - \sigma_{12} \cdot \sigma_{11}^{-1}\mu_1 + \sigma_{12} \cdot \sigma_{11}^{-1}x_1 \\ &= \mu_2 - \sigma_{12} \cdot \sigma_{11}^{-1}\mu_1 + \sigma_{12} \cdot \sigma_{11}^{-1}x_1 \\ E[X_2|X_1] &= \mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(X_1 - \mu_1) \end{aligned}$$

$$\begin{aligned} \text{Var}(X_2|X_1) &= \text{Var}(Y + \sigma_{12} \cdot \ \sigma_{11}^{-1}X_1|X_1) = \text{Var}(Y) \\ \text{Var}(Y) &= \text{Cov}(Y, X_2 - \sigma_{12} \cdot \sigma_{11}^{-1}X_1) \\ &= \text{Var}(X_2, X_2) - \sigma_{12} \cdot \ \sigma_{11}^{-1}\text{Cov}(X_1, X_2) \\ &= \sigma_{22} - \sigma_{12} \cdot \ \sigma_{11}^{-1} \cdot \ \sigma_{12} \end{aligned}$$

Mean and Variance of $X_3|X_1, X_2$

$$E(X_3|X_1, X_2) = E(X_3) + \text{Cov}(X_3, X_1|X_2)(X_1 - E(X_1|X_2)) + \text{Cov}(X_3, X_2|X_1)(X_2 - E(X_2|X_1))$$

$$\text{Var}(X_3|X_1, X_2) = \text{Var}(X_3) + \text{Cov}(X_3, X_1|X_2)^2 + \text{Cov}(X_3, X_2|X_1)^2$$

$$\text{Cov}(X_3, X_1|X_2) = \frac{\text{Cov}(X_3, X_1) - \text{Cov}(X_3, X_2) \times \text{Cov}(X_2, X_1)}{\text{Var}(X_2)}$$
$$\text{Cov}(X_3, X_2|X_1) = \frac{\text{Cov}(X_3, X_2) - \text{Cov}(X_3, X_1) \times \text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$$

$$E(X_3|X_1, X_2) = \mu_3 + \frac{\sigma_{13}}{\sigma_{11}}(X_1 - \mu_1) + \frac{\sigma_{23}}{\sigma_{22}}(X_2 - \mu_2)$$

$$\text{Var}(X_3|X_1, X_2) = \sigma_{33} - \frac{\sigma_{13}^2}{\sigma_{11}} - \frac{\sigma_{23}^2}{\sigma_{22}} - \left(\sigma_{23} - \frac{\sigma_{13} \cdot \sigma_{12}}{\sigma_{11}}\right)^2 \cdot \frac{1}{\sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}}$$

Setting seed

```r
set.seed(2024)
```

Setting mean vector

```r
mu <- c(1, 2, 3)
```

Setting covariance matrix

```r
#
variance <- 4

#
correlation <- 0.4

#
sigma <- variance^(1/2) * matrix(c(1, 0.4, 0.16,
                                   0.4, 1, 0.4,
                                   0.16, 0.4, 1) * variance^(1/2), nrow = 3)
```

Generating samples

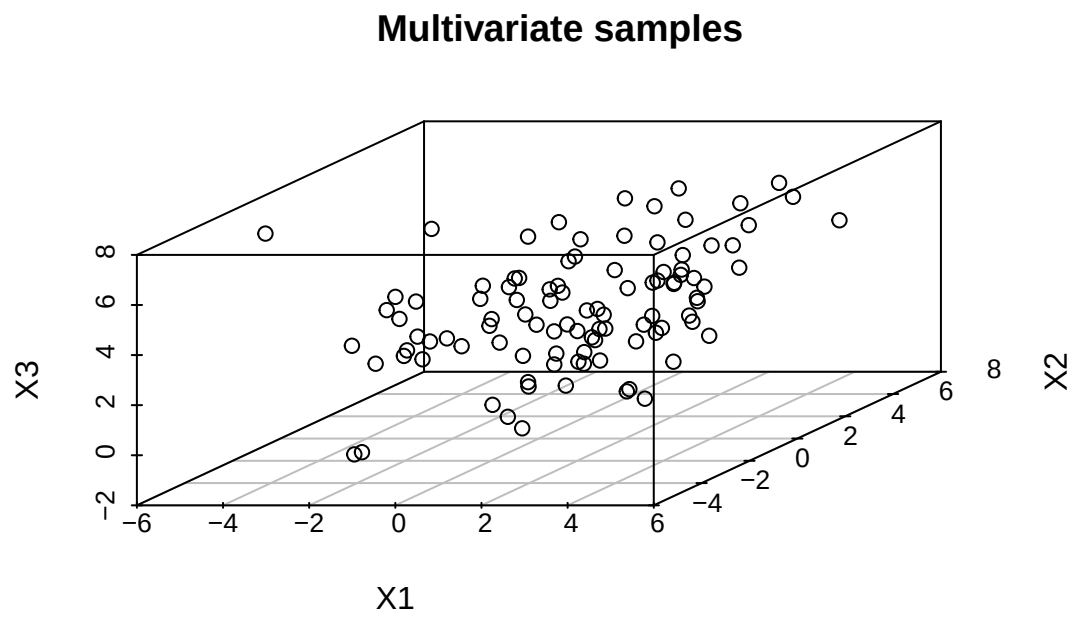
```r
n <- 100 #
# X1, X2_given_X1, X3_given_X1_X2
X1 <- rnorm(n, mean = mu[1], sd = sqrt(sigma[1, 1]))
X2 <- rnorm(n, mean = mu[2], sd = sqrt(sigma[2, 2]))
X2_given_X1 <- rnorm(n, mean = mu[2] + sigma[1,2]/sigma[1,1] * (X1 - mu[1]),
                     sd = sqrt(sigma[2, 2] - (sigma[1, 2]^2 / sigma[1, 1])))
X3_given_X1_X2 <- rnorm(n,
                        mean = mu[3] + (sigma[1, 3]/sigma[1, 1]) * (X1 - mu[1]) +
                               (sigma[2, 3]/sigma[2, 2]) * (X2 - mu[2]),
                        sd = sqrt(sigma[3, 3] - (sigma[1, 3]^2 / sigma[1, 1]) -
                                  ((sigma[2, 3] - (sigma[1, 3] * sigma[1, 2] / sigma[1, 1]))^2 /
                                  (sigma[2, 2] - (sigma[1, 2]^2 / sigma[1, 1])))))
```

Generating multivariate normal dist.

```r
multivariate_samples <- cbind(X1, X2_given_X1, X3_given_X1_X2)
```

Generating 3rd dimension scatter plot

```r
scatterplot3d::scatterplot3d(multivariate_samples[, 1], multivariate_samples[, 2], multivariate_samples
main = "Multivariate samples", xlab = "X1", ylab = "X2", zlab = "X3")
```




```r
data <- data.frame(X1, X2_given_X1, X3_given_X1_X2)
```

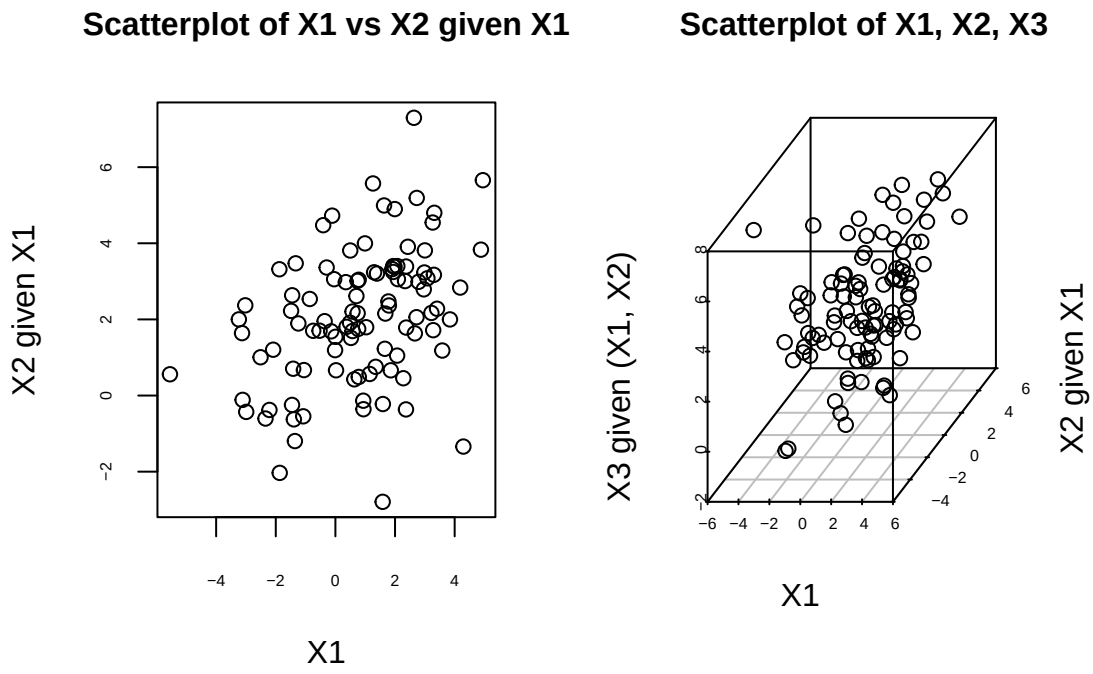
Scatter plot of $[X_1]$, $[X_2|X_1]$, Scatter plot of $[X_3|X_1, X_2]$

```r
library(scatterplot3d)

# plot
par(mfrow = c(1, 2))

#    1: X1 vs X2_given_X1
plot(X1, X2_given_X1, xlab = "X1", ylab = "X2 given X1", main = "Scatterplot of X1 vs X2 given X1", cex

#    2: X1, X2, X3
scatterplot3d(X1, X2_given_X1, X3_given_X1_X2,
              xlab = "X1", ylab = "X2 given X1", zlab = "X3 given (X1, X2)",
              main = "Scatterplot of X1, X2, X3", cex.main = 1, cex.axis = 0.5)
```




```
par(cex.lab = 0.1)
```

Comparing $[X_1]$, $[X_2|X_1]$, $[X_3|X_1, X_2]$ in histograms and graphs

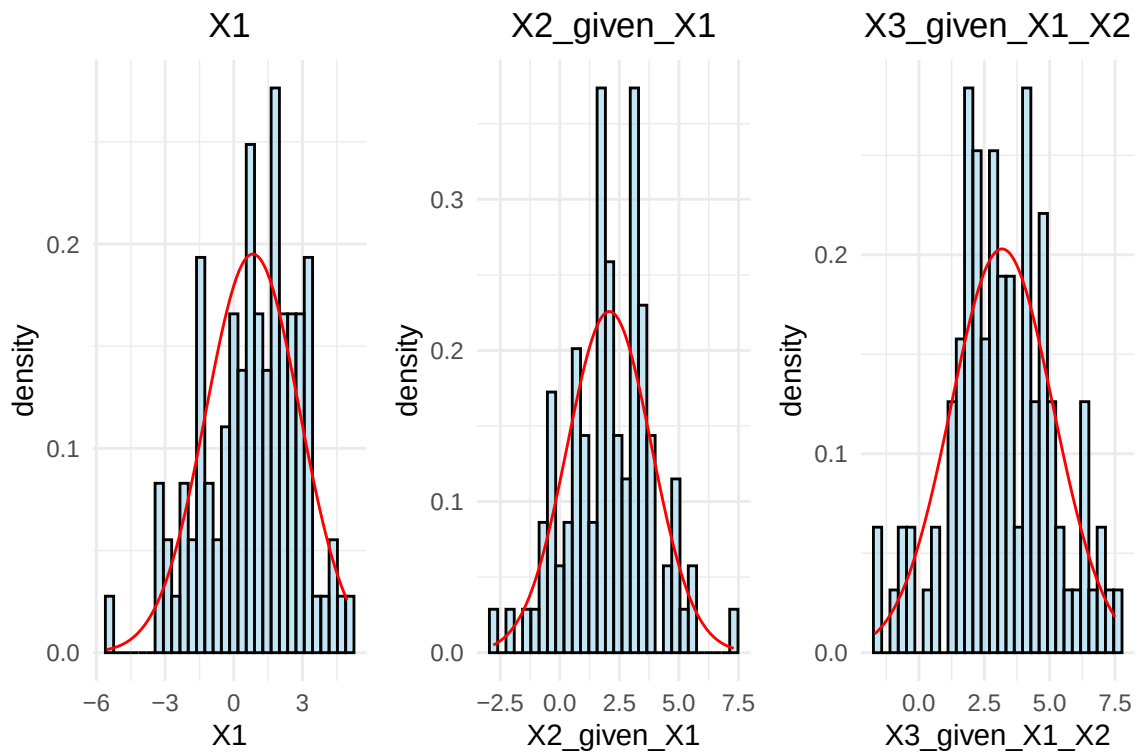
```
#
library(ggplot2)
library(gridExtra)

#
histograms <- list()

#
for (i in 1:3) {
  p <- ggplot(data, aes(x = !!sym(colnames(data)[i]))) +
    geom_histogram(aes(y = after_stat(density)), bins = 30, alpha = 0.5, color = "black", fill = "skyblu
    stat_function(fun = dnorm, args = list(mean = mean(data[[i]]), sd = sd(data[[i]])), color = "red") +
    labs(title = paste(colnames(data)[i])) +
    theme_minimal() + theme(plot.title = element_text(hjust = 0.5))

  histograms[[i]] <- p
}

#
gridExtra::grid.arrange(histograms[[1]], histograms[[2]], histograms[[3]], ncol = 3)
```



Calculating covariance

```
direct_sigma <- cov(multivariate_samples)
```

Comparing with the setting covariance

setting covariance

```
print(sigma)
```

```
##      [,1] [,2] [,3]
## [1,] 4.00  1.6 0.64
## [2,] 1.60  4.0 1.60
## [3,] 0.64  1.6 4.00
```

calculated covariance

```
round((direct_sigma),2)
```

```
##                  X1 X2_given_X1 X3_given_X1_X2
## X1             4.18        1.39           0.76
## X2_given_X1    1.39        3.12           0.96
## X3_given_X1_X2 0.76        0.96           3.86
```

Comparing with mvtnorm function

```r
library(mvtnorm)
```

```
## Warning: package 'mvtnorm' was built under R version 4.2.3
```

```r
#
set.seed(2024)

#
mu <- c(1, 2, 3)

#
sigma <- variance^(1/2)*matrix(c(1, 0.4, 0.16,
                  0.4, 1, 0.4,
                  0.16, 0.4, 1)*variance^(1/2), nrow = 3)

#
n <- 100

#
multivariate_samples <- rmvnorm(n, mean = mu, sigma = sigma)

#
round(cov(multivariate_samples), 2)
```

```
##      [,1] [,2] [,3]
## [1,] 3.61 1.95 0.41
## [2,] 1.95 4.25 1.46
## [3,] 0.41 1.46 3.96
```