Multivariate normal: the precision matrix

Matthew Stephens
University of Chicago
January 12, 2026

See here for a PDF version of this vignette.

Prerequisites

You should be familiar with the multivariate normal distribution and the idea of conditional independence, particularly as illustrated by a Markov chain.

Overview

This vignette introduces the precision matrix of a multivariate normal. It also illustrates its key property: the zeros of the precision matrix correspond to conditional independencies of the variables.

Definition, and statement of key property

Let $X = (X_1, \ldots, X_n)$ be a multivariate normal random variable with covariance matrix Σ.

The precision matrix, Ω, is simply defined to be the inverse of the covariance matrix:

$$\Omega := \Sigma^{-1}.$$

The key property of the precision matrix is that its zeros tell you about conditional independence. Specifically,

$$\Omega_{ij} = 0 \text{ if and only if } X_i \text{ and } X_j \text{ are conditionally independent given all other coordinates of } X.$$

It may help to compare this with the analogous property of the covariance matrix:

$$\Sigma_{ij} = 0 \text{ if and only if } X_i \text{ and } X_j \text{ are independent.}$$

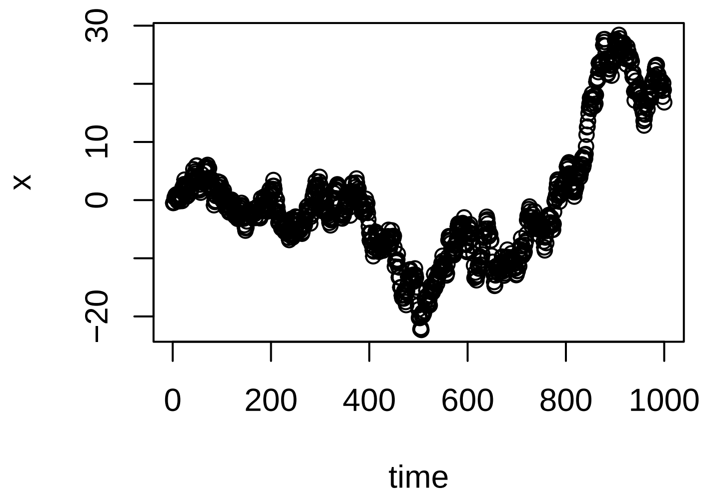
That is, whereas zeros of the covariance matrix tell you about *independence*, zeros of the precision matrix tell you about *conditional independence*.

Example: a normal Markov chain

Consider a Markov chain $X_1, X_2, \ldots$, where the transitions are given by $X_{t+1} \mid X_t \sim N(X_t, 1)$. You might think of this Markov chain as a type of "random walk": given the current state, the next state is obtained by adding a random normal value (with mean 0 and variance 1).

The following code simulates a realization of this Markov chain, starting from an initial state $X_1 \sim N(0,1)$, and plots it.

```
set.seed(100)
sim_normal_mc <- function (T = 1000) {
  x <- rep(0,T)
  x[1] <- rnorm(1)
  for (t in 2:T)
    x[t] <- x[t-1] + rnorm(1)
  return(x)
}
plot(sim_normal_mc(1000),xlab = "time",ylab = "x")
```



The normal Markov chain as a multivariate normal

If you think a little, you should be able to see that the above random walk simulation is actually simulating from a 1000-dimensional multivariate normal distribution!

Why? Well, let's write each of the $N(0,1)$ variables generated using `rnorm()` in our code as $Z_1, Z_2, \ldots$. Then we have

$$\begin{aligned} X_1 &= Z_1 \\ X_2 &= X_1 + Z_2 = Z_1 + Z_2 \\ X_3 &= X_2 + Z_3 = Z_1 + Z_2 + Z_3, \end{aligned}$$

and so on.

So we can write $X = AZ$, where A is the following 1000×1000 matrix:

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \vdots & & & & \end{pmatrix}.$$

Let's take a look at what the covariance matrix Σ looks like. (We can get a good idea from looking at the top left corner of the matrix.)

```
A <- matrix(0,1000,1000)
for(i in 1:1000)
  A[i,] <- c(rep(1,i),rep(0,1000 - i))
Sigma <- A %*% t(A)
Sigma[1:10,1:10]
#       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
#  [1,]    1    1    1    1    1    1    1    1    1     1
#  [2,]    1    2    2    2    2    2    2    2    2     2
#  [3,]    1    2    3    3    3    3    3    3    3     3
#  [4,]    1    2    3    4    4    4    4    4    4     4
#  [5,]    1    2    3    4    5    5    5    5    5     5
#  [6,]    1    2    3    4    5    6    6    6    6     6
#  [7,]    1    2    3    4    5    6    7    7    7     7
#  [8,]    1    2    3    4    5    6    7    8    8     8
#  [9,]    1    2    3    4    5    6    7    8    9     9
# [10,]    1    2    3    4    5    6    7    8    9    10
```

Now let us examine the precision matrix, Ω, which recall is the inverse of Σ. Again we just show the top left corner of the precision matrix here.

```
Omega <- chol2inv(chol(Sigma))
Omega[1:10,1:10]
#       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
#  [1,]    2   -1    0    0    0    0    0    0    0     0
#  [2,]   -1    2   -1    0    0    0    0    0    0     0
#  [3,]    0   -1    2   -1    0    0    0    0    0     0
#  [4,]    0    0   -1    2   -1    0    0    0    0     0
#  [5,]    0    0    0   -1    2   -1    0    0    0     0
#  [6,]    0    0    0    0   -1    2   -1    0    0     0
#  [7,]    0    0    0    0    0   -1    2   -1    0     0
#  [8,]    0    0    0    0    0    0   -1    2   -1     0
#  [9,]    0    0    0    0    0    0    0   -1    2    -1
# [10,]    0    0    0    0    0    0    0    0   -1     2
```

Notice all the zeros in the precision matrix. This is because of the conditional independencies that occur in a Markov chain. In a Markov chain (*any* Markov chain), the conditional distribution of X_t given the other X_s ($s \neq t$) depends only on its neighbors X_{t-1} and X_{t+1}. That is, X_t is conditionally independent of all other X_s given X_{t-1} and X_{t+1}. This is exactly what we are seeing in the precision matrix above: the non-zero elements of the tth row are at coordinates $t-1$, t and $t+1$.

Addendum: interpretation of Ω in terms of conditional mean of X_i

The following fact is also useful, both in practice and for intuition.

Suppose $X \sim N_r(0, \Omega^{-1})$, where the subscript r indicates that X is r-variate.

Let Y_1 denote the first coordinate of X, and let Y_2 denote the remaining coordinates, that is, $Y_2 := (X_2, \ldots, X_r)$. Further let Ω_{12} denote the $1 \times (r-1)$ submatrix of Ω that consists of row 1 and columns 2 through r.

The conditional distribution of $Y_1 \mid Y_2$ is (univariate) normal with mean

$$E[Y_1 \mid Y_2] = -\Omega_{12} Y_2 / \Omega_{11}$$

and variance $1/\Omega_{11}$.

Of course, there is nothing special about X_1: a similar result applies for any X_i. You just have to replace Ω_{11} with Ω_{ii} and define Ω_{12} to be the ith row of Ω with all columns except column i.

Application

An application of this is imputation of missing values: suppose one of the X values is missing, say X_i is missing, but you know the covariance matrix and all the other X values. Then you could impute X_i by its conditional mean, which is a simple linear combination of the other values that can be read directly off the ith row of the precision matrix. This idea is the essence of Kriging.

Example

Consider the Markov chain above. The conditional distribution of X_1 given all other X values is given by

$$X_1 \mid X_2, X_3, \ldots, X_{1000} \sim N(X_2/2, 1/2).$$

And the conditional distribution of X_2 given all other X values is

$$X_2 \mid X_1, X_3, X_4, \ldots, X_{1000} \sim N((X_1 + X_3)/2, 1/2).$$

And similarly for X_i, $i = 3, \ldots, 1000$. The intuition is that, if we wanted to guess what the value of X_i were given all other X's, the best guess would be the average of its neighbours.