is strictly larger than j_1. Therefore we can produce an infinite sequence of integers $j_1 < j_2 < j_3 < \dots$ such that $x \in A_{j_i}$ for all i. Let E be the event $\{x : x \in A_i, \text{i.o.}\}$. By the definition of $\limsup_{n\to\infty} A_n$ we have

$$E = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$$

From $E \subset B_k$ for all k, it follows that $P(E) \le P(B_k)$ for all k. By the property of union bound, we have that $P(B_k) \le \sum_{i=k}^{\infty} P(A_i)$. By hypothesis since $\sum_{i=1}^{\infty} P(A_i)$ is finite and hence $P(B_k) \to 0$ as $k \to \infty$. Therefore $P(E) = 0$. ∎

Proof (Borel-Cantelli 2) Let E denote the set of samples that are in A_i infinitely often. We have to show that complement of E (denoted by E^c) has probability zero.

Taking the complement of E we find using the definition in the proof of Borel-Cantelli 1) and DeMorgan's laws

$$E^c = \bigcup_{k=1}^{\infty} \bigcap_{i=k}^{\infty} A_i^c$$

But for each k, assuming that the A_i's are independent,

$$\begin{aligned} P\big(\bigcup_{i=k}^{\infty} A_i^c\big) &= \prod_{i=k}^{\infty} P(A_i^c) \\ &= \prod_{i=k}^{\infty} (1 - P(A_i)) \end{aligned}$$

The inequality $1 - a \le e^{-a}$ and the assumption that the sum of $P(A_i)$ diverges together imply that

$$P\big(\bigcup_{i=k}^{\infty} A_i^c\big) \le \exp\Big(-\sum_{i=k}^{\infty} P(A_i)\Big) = 0$$

Therefore E^c is a union of countable number of events, each of them has probability zero. So $P(E^c) = 0$. ∎

14.13 Partitioned Matrices

Consider the matrix given by

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where A and D are square matrices not necessarily of the same size and B and C are not necessarily square matrices. Assume A and D have inverses A^{-1} and D^{-1} respectively. Multiply the top row by CA^{-1} and subtract that result from the bottom row. The resulting matrix is

$$M' = \begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix}$$

Now this type of transformation can be thought of as a particular type of elementary row operation that leaves the determinant of M unchanged: $\det M = \det M'$. Using the notation $|M|$ as a shorthand for $\det M$ we can conclude

$$|M| = |A| \cdot |D - CA^{-1}B|.$$

By doing a similar row operation by multiplying the second row by BD^{-1} and subtracting it from the top row we obtain

$$M'' = \begin{bmatrix} A - BD^{-1}C & 0 \\ C & D \end{bmatrix}$$

So we also have that $\det M = \det M''$ and hence

$$|M| = |D| \cdot |A - BD^{-1}C|.$$

The above identities are handy in dealing with some of the applications of stochastic processes to the problem of filtering.

Let us develop a couple of interesting matrix identities which occur in filters. We will develop these identities by performing the inversion procedure on the matrix M using the following work matrix to keep track of the inversion steps:

$$\left[\begin{array}{cc|cc} A & B & I & 0 \\ C & D & 0 & I \end{array}\right],$$

First subtract CA^{-1} times the first row from the second row and multiply the first row by A^{-1}. The work matrix becomes:

$$\left[\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & D - CA^{-1}B & -CA^{-1} & 1 \end{array}\right],$$

Next subtract $A^{-1}B[D - CA^{-1}B]^{-1}$ times the second row from the first row and multiply the second row by $[D - CA^{-1}B]^{-1}$. The work matrix now becomes:

$$\left[\begin{array}{cc|ll} I & 0 & A^{-1} - A^{-1}B[D - CA^{-1}B]^{-1}CA^{-1} & -A^{-1}B[D - CA^{-1}B]^{-1} \\ 0 & I & -[D - CA^{-1}B]^{-1}CA^{-1} & [D - CA^{-1}B]^{-1} \end{array}\right],$$

Now, in order to discover the interesting identities let use an alternative ordering of row operations. In this second case we first subtract BD^{-1} times the second row from the first row and multiply the second row by D^{-1}. The work matrix becomes:

$$\left[\begin{array}{cc|cc} A - BD^{-1}C & 0 & I & -BD^{-1} \\ D^{-1}C & I & 0 & D^{-1} \end{array}\right],$$

Next subtract $D^{-1}C[A-BD^{-1}C]^{-1}$ times the first row from the second row and multiply the first row by $[A-BD^{-1}C]^{-1}$. The work matrix now becomes:

$$\left[\begin{array}{cc|cc} I & 0 & [A-BD^{-1}C]^{-1} & -[A-BD^{-1}C]^{-1}BD^{-1} \\ 0 & I & -D^{-1}C[A-BD^{-1}C]^{-1} & D^{-1}-D^{-1}C[A-BD^{-1}C]^{-1}BD{-1} \end{array}\right],$$

Since these are two ways of obtaining the inverse of M, the matrix elements must be equal term-by-term. Hence we obtain some interesting identities. For example:

$$[A-BD^{-1}C]^{-1} = A^{-1} - A^{-1}B[D-CA^{-1}B]^{-1}CA^{-1}. \tag{71}$$

Let us apply the ideas of joint random variables, conditional probabillity and expectation to a random $(n+m)$-vector $Z^T = [X^T, Y^T]$, where X is an n-vector and Y is an m-vector and assume that Z is a multivariate Gaussian $Z \sim N(m_Z, P_{ZZ})$ (with $m_Z = E[Z]$ is a vector of means and $P_{ZZ} = E[(z-m_Z)(z-m_Z)^T]$ is the covariance matrix). This also assumes that X and Y are jointly random Gaussian variables. Then Z has a probability density function given by

$$f_Z(z) = \frac{1}{\sqrt{(2\pi)^{n+m}|P_{ZZ}|}} \exp\left\{-\frac{1}{2}(z-m_z)^T P_{ZZ}^{-1}(z-m_z)\right\}. \tag{72}$$

Now we can use the partition matrix approach given in Section [14.13] to write the covariance matrix P_{ZZ} as

$$P_{ZZ} = E[(z-m_Z)(z-m_Z)^T] = \begin{bmatrix} E[(X-m_X)(X-m_X)^T] & E[(X-m_X)(Y-m_Y)^T] \\ E[(Y-m_Y)(X-m_Z)^T] & E[(Y-m_Y)(Y-m_Y)^Y] \end{bmatrix} = \begin{bmatrix} P_{XX} & P_{XY} \\ P_{YX} & P_{YY} \end{bmatrix}$$

Then we have

Theorem 14.14 *If the random vectors X and Y are distributed according to Eq. [72] then X and/or Y is also a multivariate Gaussian (with appropriate parameters).*

Proof Let show the theorem holds for Y. Accordingly let use consider a transformation of variables to $W^T = [W_1^T, W_2^T]$ such that

$$W_1 = X - P_{XY}P_{YY}^{-1}Y, \quad W_2 = Y.$$

Observe that the Jacobian of the above transformation is 1.

From Theorem [5.1] W is normally distributed. Let us compute $E[(W_1 - E[W_1])(W_2 - E[W_2])]$,

$$\begin{aligned} E[W_1] &= m_X - P_{XY}P_{YY}^{-1}m_Y \\ W_1 - E[W_1] &= (X-m_X) - P_{XY}P_{YY}^{-1}(Y-m_Y) \\ E[W_2] &= m_Y \\ W_2 - E[W_2] &= (Y-m_Y) \end{aligned}$$