Vector space

- $m \times n$ 행렬 A를 m차원 열벡터 n개의 모음 $[\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$ 으로 볼 때, $A\mathbf{x} = \mathbf{b}$ 를 $\mathbf{x} = [x_1, x_2, \ldots, x_n]'$ 에 대해 $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n = \mathbf{b}$ 으로 볼 수 있음, 이 경우 $\mathbf{b}$ 가 A의 열벡터들의 선형결합, 즉 A의 column space의 원소여야만 nontrivial solution $\mathbf{x}$ 가 존재함.

- $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ 이 선형독립인 벡터들이면 열공간의 차원(즉 rank) 은 n이지만, 선형종속이면 열공간의 차원은 n보다 낮음. 선형종속인 경우의 열공간의 차원이 r일 때, 즉 선형독립인 열벡터가 r개일 때는 r개의 서로 독립인 nontrivial solution $\mathbf{x}$들을 얻음. 이 경우 A의 영공간 ($A\mathbf{x} = \mathbf{0}$ 인 $\mathbf{x}$의 집합) 의 차원은 $n - r$ 임.

- 행렬의 행공간의 rank와 열공간의 rank는 같음. 행렬을 r차원 공간으로 보내는 사상으로 이해하면 자명함. 같은 직관으로 $A'A$의 rank와 AA'의 rank는 A의 rank와 같고, AB의 rank와 BA의 rank는 A 또는 B의 rank 보다 클 수 없음.

- 다만 행공간과 열공간이 개념적으로 완전히 같은 것은 아님. 만약 $m \times n$ 행렬에서 $m > n$ 이고 rank가 n인 경우, 행공간은 n차원 공간 전체이지만, 열공간은 m차원 공간 안에서 n차원을 갖는 부분공간임.

- A의 열공간 $A\mathbf{y}$는 A의 좌영공간 $\{\mathbf{x}|A'\mathbf{x} = \mathbf{0}\}$ 과 직교함 ($\mathbf{x}'A\mathbf{y} = (A'\mathbf{x})'\mathbf{y} = \mathbf{0}'\mathbf{y} = 0$). 마찬가지로 A의 행공간과 A의 영공간은 직교함.

- $n \times n$ 정사각행렬 A가 역행렬 A^{-1}을 가진다는 것은 A의 열공간의 차원이 n, 즉 열벡터들이 선형독립임을 의미함. 이 경우 $A\mathbf{x} = \mathbf{b}$ 의 해 $\mathbf{x}$는 유일함.

- $C = AB$에서 B의 j번째 열의 원소들은, C의 j번째 열을 구성하는 데 A의 각 열벡터들이 몇 개씩 쓰였는지를 보여줌. 한편 B의 i번째 행의 원소들은, C의 각 열벡터 별로 A의 i번째 열벡터가 몇 개씩 쓰였는지를 보여줌.

- 위 관계를 이용해, 행렬곱 AB를 A의 열벡터 $\mathbf{a}_j$와 B의 행벡터 $\mathbf{b}_j^*$ 간의 곱인 rank가 1인 행렬들의 합으로, 즉 $AB = \mathbf{a}_1\mathbf{b}_1^* + \mathbf{a}_2\mathbf{b}_2^* + \ldots + \mathbf{a}_n\mathbf{b}_n^*$ 으로 나타낼 수 있음. AB의 각 열벡터 구성에 $\mathbf{a}_j$가 몇 번씩 쓰였는지를 $\mathbf{b}_j^*$를 통해 알 수 있음. 이를테면 $\mathbf{b}_1^* = [b_{11}\, b_{12}]$ 이면 AB의 1열에는 $\mathbf{a}_1$이 b_{11}개, 2열에는 $\mathbf{a}_1$이 b_{12}개 쓰임.

- 이러한 rank 1인 조각들의 합으로 나타내는 것을 응용한 예가 LU 분해임, 피보팅 연산 및 $A = LU = \mathbf{l}_1\mathbf{u}_1^* + \mathbf{l}_2\mathbf{u}_2^* + \ldots + \mathbf{l}_n\mathbf{u}_n^*$ 를 통해, 대각성분이 전부 1인 하삼각행렬 L, 그리고 상삼각행렬 U의 곱으로 분해할 수 있음.

- LU분해 시 $A\mathbf{x} = \mathbf{b}$ 를 푸는 것은 $L\mathbf{y} = \mathbf{b}$를 풀어 얻은 $\mathbf{y}$ 를 갖고 $U\mathbf{x} = \mathbf{y}$ 를 푸는 것이 되며, L과 U 각각에 대한 문제들은 삼각행렬이라는 구조적 특성상 단순 대입법만으로 풀리므로 연산량이 약 $2n^3/3$에서 $2n^2$ 정도로 줄어듦. 따라서 같은 A에 대해 여러 가지 $\mathbf{b}$를 써서 문제를 풀어야 할 경우 A를 한 번 LU분해해 두면 유용함.

- 가끔 LU분해의 피보팅 과정에서 0으로 나누는 경우가 생길 수 있는데, 이러한 문제를 방지하기 위해 치환행렬 (어떤 행에도 1은 하나씩, 어떤 열에도 1은 하나씩 있어 성분의 순서를 재정렬하는 행렬) P를 곱하면 $PLU\mathbf{x} = \mathbf{y}$ 이고, P는 직교행렬이므로 $P^{-1} = P'$를 이용하면 $LU\mathbf{x} = P'\mathbf{y}$ 가 됨.

- $m \times n$ 행렬 A의 열벡터들 중 선형독립인 r개의 벡터들로 이루어진 subset만 모아서 $m \times r$ 행렬 C에 집어넣을 때 C의 열벡터들을 기저(basis)로 볼 수 있으며, 각 기저가 몇 개씩 쓰였는지를 나타내는 $r \times n$ 행렬 R을 도입해 $A = CR$ 로 분해 가능함.

- 계산의 간편화를 위해 기저벡터들이 서로 직교하고 길이가 1이 되기를 (즉 정규직교기저가 되기를) 희망하는데, 서로 독립인 벡터들로부터 정규직교기저를 생성하는 Gram-Schmidt 방법이 있음.

  먼저 $\mathbf{x}_1$ 방향의 단위벡터 $\mathbf{e}_1$를 결정함. 그리고 $\mathbf{x}_2$를 $\mathbf{e}_1$에 정사영시킨 벡터를 $\mathbf{x}_2$로부터 빼면 $\mathbf{e}_1$과 직교하는 벡터가 되며 그것의 길이로 나누어 $\mathbf{e}_2$를 얻음.

  그리고 $\mathbf{x}_3$을 $\mathbf{e}_1, \mathbf{e}_2$ 각각에 정사영시킨 벡터들의 합을 $\mathbf{x}_3$으로부터 빼면 $\mathbf{e}_1, \mathbf{e}_2$ 모두와 직교하는 벡터가 되며 그것의 길이로 나누어 $\mathbf{e}_3$를 얻음. 이를 반복해 정규직교기저들을 얻음.

Eigenvectors, eigenvalues, and eigen decomposition

- $n \times n$ '정사각'행렬에 대해 $A\mathbf{x} = \lambda\mathbf{x}$ 를 만족시키는 수 λ와 벡터 $\mathbf{x} \neq \mathbf{0}$를 각각 고유값, 고유벡터라 함. A의 고유벡터는 선형변환 후의 벡터 $A\mathbf{x}$가 $\mathbf{x}$와 나란함을 의미함. 즉 해당 방향의 벡터는 선형변환 A에 의해 방향은 바뀌지 않고 신축만 됨 (음의 신축률도 포함할 경우).

- 고유값은 0일 수도 있음에 주의. 고유값 0을 갖는 고유벡터가 존재한다는 것은 영공간의 차원이 0보다 크다는 것, 그리고 A의 역행렬이 존재하지 않는다는 것을 의미함.

- 서로 독립인 n개의 고유벡터가 존재하면, 각 고유벡터를 열벡터로 하는 가역행렬 P와 각 고유벡터에 대응되는 고유값을 대각성분들로 갖는 대각행렬 Λ를 써서 $A = P\Lambda P^{-1}$ 로 대각화, 고유값 분해가 가능함. 만약 서로 독립인 고유벡터의 수가 n보다 적으면, P^{-1}이 존재하지 않아 대각화가 불가능함. 이 때 P가 가역이라고 반드시 A도 가역인 것은 아님에 주의. P가 가역이면서 0인 고유값이 없어야 A도 가역임. 또한 A가 가역이어도 대각화가 불가능할 수도 있음.

- 행렬 A가 대각화 가능하다면, A^k를 직접 계산할 필요 없이 $A = P\Lambda^k P^{-1}$ 로 쉽게 계산할 수 있음, Λ가 대각행렬이기 때문임.

- 만약 A가 n개의 서로 다른 고유값을 가지면 P는 반드시 가역행렬임, 서로 다른 방향의 고유벡터가 n개 있는 것이기 때문. 그러나 서로 다른 고유값의 개수가 n개가 되지 않더라도 서로 독립인 고유벡터가 n개 존재할 수도 있음 (고유값을 신축률의 관점에서 보면 방향이 다른 두 개 이상의 고유벡터들의 신축률이 우연히 같을 수도 있으므로). 물론 서로 다른 고유값의 개수가 n개가 되지 않으며 서로 독립인 고유벡터도 n개가 되지 않을 수도 있음.

- 고유값 분해가 가능한 선형변환 A에 대한 $A\mathbf{x}$는, 표준직교기저 기준 좌표 $\mathbf{x}$에서 고유벡터기저 P 기준의 좌표 $\mathbf{y} = P^{-1}\mathbf{x}$로 좌표변환해 $\mathbf{x}$가 A의 각 고유벡터의 몇 배씩의 합인지 구한 후, 그 $\mathbf{y}$에 Λ를 곱해 각 고유벡터 축별 신축률에 따라 신축하여 변환을 수행한 후, 그 결과인 Λy 를 다시 표준직교기저 기준 좌표가 되게 $P\Lambda\mathbf{y}$ 로 역좌표변환한 결과로 볼 수 있음.

  (위에서 좌표변환이 P^{-1}인 이유는, $\mathbf{x}$를 P의 열벡터의 결합으로 나타낼 경우 $y_1\mathbf{p}_1 + y_2\mathbf{p}_2 + \ldots$ 이므로 $\mathbf{x} = P\mathbf{y}$, 그런데 우리가 구하려는 것은 P의 열벡터가 기저인 공간에서의 $\mathbf{x}$의 좌표 $y_1, y_2, \ldots$ 이므로 $\mathbf{y} = P^{-1}\mathbf{x}$)

- 행렬 A의 고유벡터들이 직교하기 위한 필요충분조건은 $A'A = AA'$인 것임, 아래서 언급할 대칭행렬은 이 조건을 만족하므로 고유벡터들이 직교함.

- (실수) 대칭행렬 ($A' = A$)인 경우 모든 고유값은 실수이며, '서로 다른' 고유값들에 대응하는 고유벡터들은 모두 직교하고, 어떤 고유값이 특성방정식의 k중근 (대수적 중복도 k) 인 경우 해당 고유값에 대해 선형독립인 고유벡터 k개를 얻을 수 있으므로, 모든 대칭행렬은 항상 직교대각화 가능함.

- 이를 종합하면, 대칭행렬은 서로 직교하는 단위 고유벡터들로 이루어진 직교행렬 Q를 이용해 $A = QVQ'$로 대각화가 가능함. Q가 정규직교기저를 제공할 뿐 아니라, Q^{-1}을 역행렬 연산으로 직접 구할 필요 없이 Q'를 쓰면 되므로 계산이 효율적임. 실제 활용에선 데이터의 공분산행렬 등이 대칭행렬이므로 이런 이점들을 누릴 수 있음.

- 위에서 각 대각성분이 각 고유값의 양의 제곱근인 행렬 $D^{1/2}$와 $B = D^{1/2}Q'$ 를 생각하면 대칭행렬 $A = B'B$로 분해한 것으로 볼 수 있음.

- 한편 대칭행렬의 고유값들의 합은 대각성분들의 합(trace)과 같고, 고유값들의 곱은 determinant와 같음.

- 만약 어떤 '대칭'행렬 S에 대해, 영벡터가 아닌 모든 벡터 $\mathbf{x}$에 대해 $\mathbf{x}'S\mathbf{x} > 0$ 이 성립한다면 S를 양의 정부호라 함. 이는 선형변환 S 전의 벡터 $\mathbf{x}$와 후의 벡터 $S\mathbf{x}$가 항상 예각을 이룸을 의미하므로, 모든 고유값이 양수여야 하며 그 역도 성립 (하나라도 음수면 특정 $\mathbf{x}$에 대해서는 $\mathbf{x}$와 Sx가 둔각을 이루게 될 수도 있음).

- S가 양의 정부호라면 $\mathbf{x}$가 영벡터가 아닌 한 $S\mathbf{x} \neq \mathbf{0}$이므로, 양의 정부호이면 full rank임. 대우를 생각하면, 대칭행렬 A가 full rank가 아닌 경우 $A\mathbf{x} = \mathbf{0}$일 수 있으므로 A는 양의 정부호가 아님. 역을 생각하면, A가 대칭이고 full rank이더라도 양의 정부호가 아닐 수 있음 (당장 음의 정부호일 수도 있고).

- 데이터 행렬처럼 열들이 독립인 행렬 A에 대해 $S = A'A$인 S는 양의 정부호임 (A가 full rank인 경우 $A'A$도 full rank이며 영벡터가 아닌 모든 $\mathbf{x}$에 대해 $\mathbf{x}'A'A\mathbf{x} = (A\mathbf{x})'(A\mathbf{x}) > 0$ 이기 때문). 그러므로 데이터의 공분산행렬은 양의 정부호임.

- 양의 정부호 행렬 S는 반드시 어떤 정방 상삼각행렬 U에 대해 $S = U'U$ 로, 또는 어떤 정방 하삼각행렬 L에 대해 $S = LL'$ 로 분해할 수 있음 (Cholesky Factorization). 또한 양의 정부호 행렬 S에 대해, $B^2 = S$인 (즉 전치 없이 진짜 제곱해서 S가 나오는) 양의 정부호 행렬 B가 유일하게 존재함.

- 다변수함수의 미분에서 Gradient가 영벡터이면서 Hessian이 양의 정부호 행렬이 되는 point가 함수를 최소로 만드는 point이고, 반대로 Hessian이 음의 정부호 행렬이 되는 point가 함수를 최대로 만드는 point이며, 만약 Hessian이 양의 정부호도 음의 정부호도 아니면 안장점임.

- 고유값을 구하는 것은 기본적으로 계산비용이 큼. 이 때 대각행렬이거나 상삼각 또는 하삼각행렬인 경우 고유값은 대각성분들 그 자체라는 좋은 성질을 이용하기 위해, 큰 행렬을 닮음변환을 반복해 행렬을 삼각행렬에 근접시키는 Jacobi법 또는 QR법을 사용함. 직접 특성방정식을 풀려고 한다면 5차 이상만 되어도 정해진 절차로 해를 얻을 수 없기 때문.

- 닮음변환의 컨셉은, 대각화가능한 A와 가역인 B에 대해 $(BAB^{-1})(B\mathbf{x}) = BA\mathbf{x} = B\lambda\mathbf{x} = \lambda(B\mathbf{x})$ 가 되어 A의 고유값인 λ가 BAB^{-1}의 고유값이기도 하므로, BAB^{-1}가 삼각행렬이 되도록 적절한 B를 곱해나가는 변환임.

- 추이행렬 (또는 확률행렬) A는 A_{ij}가 상태 j에서 상태 i로 이동할 확률이고 (그러므로 각 열벡터가 모든 성분들이 0 이상 1 이하의 값을 가지며 모든 성분의 합이 1임), 이는 한 상태에서 다른 상태로 바뀌는 확률이 오직 그 두 상태에 의해서만 결정되며 상태의 수가 유한한 Markov Chain을 나타낼 수 있음.
  $\mathbf{p}$를 각 상태별 상대비율을 담은 분포벡터라 하면, 상태가 바뀜에 따라 분포벡터가 $A\mathbf{p}, A^2\mathbf{p}, A^3\mathbf{p}, \ldots$ 로 바뀌고, 궁극적으로 어떤 극한분포 $\mathbf{p}_f$ 에 도달하면 더 이상 해당 Markov process로는 분포벡터가 바뀌지 않게 됨 ($A\mathbf{p}_f = \mathbf{p}_f$).
  이 때 식 $A\mathbf{p}_f = \mathbf{p}_f$ 을 보면 $\mathbf{p}_f$ 는 A의 고유값 1에 대한 고유벡터임. 즉 추이행렬의 고유값 1에 대한 고유벡터를 찾아서 극한분포를 구할 수 있음 (이 극한분포는 유일하게 존재함).

Singular value decomposition and pseudoinverse

- $m \times n$ 행렬 A (rank는 r) 가 정사각행렬이 아닌 경우라도 분해가 가능함, 구체적으로는 $AV = U\Sigma$ ($A = U\Sigma V'$) 로 분해됨.

- 위에서 V는 각 열이 양의 정부호 행렬 $A'A$의 정규직교 고유벡터인 $n \times n$ 직교행렬, Σ는 r번째까지의 대각성분들이 $A'A$의 고유값의 양의 제곱근들의 내림차순이고 A의 rank r번째를 넘어간 대각성분 및 모든 비대각성분이 0인 $m \times n$ 직사각 희소행렬, U는 m차원 공간의 정규직교기저들을 열벡터로 갖는 정사각행렬임. (결과적으로 보면 U의 각 열은 AA'의 정규직교 고유벡터들임)

- Σ의 양수 대각성분들 $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$ 들을 특이값이라 하며, 특이값들은 A에 의해 유일하게 결정됨.

- U의 $j(\leq r)$번째 열벡터 $\mathbf{u}_j$ 에 대해, $A\mathbf{v}_j = \sigma_j\mathbf{u}_j$ 가 성립하므로 $\mathbf{u}_j$를 찾을 수 있음. 한편 $j > r$인 경우에는 $A\mathbf{v}_j = 0$ 임.

- 특이값 분해의 또다른 의미는, 서로 직교하는 열벡터들을 행렬 A를 통해 선형변환 시 크기는 (특이값들을 factor로 가지며) 변하지만 선형변환 결과 벡터들이 여전히 서로 직교하게 되는 벡터들의 모음 V, U를 찾는 것임.

- 특이값 분해는 파일 압축에 사용됨. $m > n$이고 rank가 n인 데이터 행렬을 특이값 분해 후 만약 큰 순서대로 $k < n$개의 특이값 및 이에 대응되는 좌측부터 k개의 V 및 U의 열벡터들만 쓴다면, U_{adj}는 $m \times k$, Σ_{adj}는 $k \times k$, V'_{adj}는 $k \times n$이 되어 $A_{adj} = U_{adj}\Sigma_{adj}V'_{adj}$ 가 되고, k가 적당한 값인 경우 A_{adj}는 A와 매우 유사하면서도 필요한 행렬의 크기는 상당히 줄어들게 됨. (이게 Principal Component Analysis의 활용임)

- Rayleigh quotient $\|A\mathbf{x}\|^2/\|\mathbf{x}\|^2 = (\mathbf{x}'A'A\mathbf{x})/(\mathbf{x}'\mathbf{x})$의 최대값을 찾는 문제 (A를 타원 변환으로 볼 때 장축을 찾는 문제 등) 에서, 최대값은 $A'A$의 가장 큰 고유값, 즉 A의 가장 큰 특이값의 제곱 σ_1^2 이며 최적의 $\mathbf{x}$는 그 고유값에 대응되는 $A'A$의 고유벡터 $\mathbf{v}_1$임.

- Rayleigh quotient 최대값 문제를 $\mathbf{v}_1'\mathbf{x} = 0$ 조건 하에서 풀 경우 최대값은 $A'A$의 두 번째로 큰 고유값, 즉 A의 두 번째로 큰 특이값의 제곱 σ_2^2 이며 최적의 $\mathbf{x}$는 그 고유값에 대응되는 $A'A$의 고유벡터 $\mathbf{v}_2$임. 한편 Rayleigh quotient의 최소값은 A의 가장 작은 특이값 σ_r^2이며 이 때의 $\mathbf{x}$는 그 특이값에 대응되는 $A'A$의 고유벡터 $\mathbf{v}_r$임.

- 일반화된 Rayleigh quotient $(\mathbf{x}'S\mathbf{x})/(\mathbf{x}'M\mathbf{x})$ 의 경우 최대값은 $M^{-1}S$ 의 고유값 중 가장 큰 값임. 그러나 $M^{-1}S$은 대칭이 아닐 수도 있으므로, 이와 같은 고유값을 갖는 대칭행렬 $H = M^{-1/2}SM^{-1/2}$ 을 이용함. 이 경우 $\mathbf{x} = M^{-1/2}\mathbf{y}$ 라 하면, 일반화된 레일리 몫은 $(\mathbf{y}'H\mathbf{y})/(\mathbf{y}'\mathbf{y})$ 가 되어, 레일리 몫이 됨.

- 일반화된 레일리 몫은 일반화된 고유벡터와 고유값, $S\mathbf{x} = \lambda M\mathbf{x}$ 를 만족하는 $\mathbf{x}$ 와 λ 찾기로 볼 수 있음, 이 때 S와 M이 양의 정부호라면, 이 식을 만족하는 서로 다른 고유값들에 대한 두 벡터 $\mathbf{x}_1$ 과 $\mathbf{x}_2$ 간에는 직교는 아니지만 M-직교, 즉 $\mathbf{x}_1'M\mathbf{x}_2 = 0$ 이 성립함.

- 열 개수가 같은 두 행렬 A와 B를 각각 특이값 분해할 때, 동일한 가역행렬(직교는 아닐 수 있음) Z에 대해 $A = U_A\Sigma_AZ$, $B = U_B\Sigma_BZ$로 분해할 수 있음. 이 경우 U_A와 U_B는 직교행렬이고, $\Sigma_A'\Sigma_A + \Sigma_B'\Sigma_B = I_{n\times n}$ 이 성립함.

- Fisher's LDA에서 분리 비율 $R = (\mathbf{x}'\mathbf{m}_1 - \mathbf{x}'\mathbf{m}_2)^2/(\mathbf{x}'\Sigma_1\mathbf{x} + \mathbf{x}'\Sigma_2\mathbf{x})$ 를 최대화하는 문제는 $S = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)', M = \Sigma_1 + \Sigma_2$로 둘 경우 일반화된 레일리 몫 $(\mathbf{x}'S\mathbf{x})/(\mathbf{x}'M\mathbf{x})$ 의 최대화 문제가 됨, 이 때 S의 rank가 1이므로 S의 열공간은 $\mathbf{m}_1 - \mathbf{m}_2$ 방향의 직선이고 $S\mathbf{v} = \lambda M\mathbf{v}$으로부터 $M\mathbf{v}$도 같은 방향이므로, R을 최대화하는 벡터는 $\mathbf{v} = M^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$ 가 됨.

- 직사각행렬 A의 열들이 독립인 경우 (일반적으로 아래로 긴 데이터행렬) 유사역행렬 $A^+ = (A'A)^{-1}A'$ 로 정의됨. $A^+A = I$이기 때문. 만약 행들이 독립인 경우라면 $A^+ = A'(AA')^{-1}$ 임. 또한 A를 특이값 분해된 $U\Sigma V'$ 로 나타낼 경우, $A^+ = V\Sigma^+U'$가 되고, 여기서 Σ^+는 Σ를 전치 후 대각성분들의 역수를 취한 행렬임.

- $A\mathbf{x} = \mathbf{b}$의 정확한 해가 없을 경우, $\|A\mathbf{x} - \mathbf{b}\|$ 를 최소화하는, 즉 least square의 solution $\hat{\mathbf{x}} = A^+\mathbf{b}$ 임. (A는 설명변수들 데이터 행렬, $\mathbf{b}$는 반응변수 벡터, $\mathbf{x}$는 계수추정량)

- 반응변수의 실제값 벡터 $\mathbf{b}$를 A의 열공간에 정사영한 점은 $A\hat{\mathbf{x}} = A(A'A)^{-1}A'b$ 임, 즉 b를 A의 열공간으로 정사영시키는 행렬은 $A(A'A)^{-1}A'$, 회귀분석에서 hat matrix로 불리는 행렬임.

Probability and statistics

- 이변량 이산확률변수에서 첫 번째 확률변수를 x_i, 두 번째 확률변수를 y_j라 할 때, 각 사건의 확률 p_{ij} 및 각 변수의 평균들 m_x, m_y가 알려진 경우, 공분산행렬은 아래와 같이 계산됨.

$$V = \sum_i \sum_j p_ij \begin{bmatrix} (x_i - m_x)^2 & (x_i - m_x)(y_j - m_y) \\ (x_i - m_x)(y_j - m_y) & (y_j - m_y)^2 \end{bmatrix}$$

- 각 확률변수끼리 완전한 선형종속이 아닌 한, 공분산행렬은 양의 정부호임.

- M차원 공분산행렬을 V라 할 때 각 대각성분은 $\sigma_1^2, \ldots, \sigma_M^2$ 임. 이 때 대각행렬 $D = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_M^{-1})$ 이라 하면 (즉 각 변수 별 분산의 역수의 제곱근을 대각성분으로 하는 대각행렬), 상관계수행렬은 $R = DVD$로 계산됨.

- 다변량확률변수 X를 두 개의 subset X_a, X_b로 나누고 공분산행렬 V를 4개의 block들 V_{aa}, V_{ab}, V_{ba}, V_{bb} 을 묶은 형태로 볼 경우, X_b에 conditional한 X_a만의 공분산 $V_{a|b}$ 는 $V_{aa} - V_{ab}V_{bb}^{-1}V_{ba}$ 로 계산됨.

- 다변량정규분포의 중적분을 계산할 때 공분산이 있는 경우, 대각화를 이용해 서로 독립인 변수를 새롭게 만들면 보다 쉽게 계산 가능함.

- 음이 아닌 스칼라 확률변수에 대한 마르코프 부등식 $P(X \geq a) \leq E[X]/a$ 을 확률변수 행렬로 확장하면, 양의 준정부호인 확률변수 행렬 X 및 어떤 양의 정부호 행렬 A에 대해, $A - X$가 양의 준정부호가 아닐 확률의 상한은 $E[X]A^{-1}$의 대각합임.

- 스칼라 확률변수에 대한 체비셰프 부등식 $P(|X - E[X] \geq a) \leq \sigma^2/a^2$ 을 다변수확률변수로 확장하면, M 차원 확률변수 벡터 X, 공분산행렬 V 및 어떤 주어진 스칼라 a에 대해, $\sqrt{(X - E[X])'V^{-1}(X - E[X])}$ 가 a 보다 클 확률의 상한은 M/a^2 임.

Perturbation in matrices

- 단위행렬로부터 약간의 perturbation이 생긴 $n \times n$ 행렬 $M = I - \mathbf{u}\mathbf{v}'$ 의 역행렬은 $M^{-1} = I + (\mathbf{u}\mathbf{v}')/(1 - \mathbf{v}'\mathbf{u}$ 로 주어짐. ($\mathbf{u}\mathbf{v}'$는 rank가 1임).

- 위에서 rank 1인 행렬 $\mathbf{u}\mathbf{v}'$ 를 rank k인 행렬 UV'로 확장하면, $M^{-1} = I_n + U(I_k - V'U)^{-1}V'$ 로 주어짐. 여기서 역행렬 계산을 크기 n인 행렬 대신 크기 k인 행렬에 대해 해도 M^{-1}을 구할 수 있음.

- 위를 일반화해서 $M = A - UV'$인 경우 $M^{-1} = A^{-1} + A^{-1}U(I_k - V'A^{-1}U)^{-1}V'A^{-1}$ 임, 즉 A^{-1} 을 알거나 계산이 쉬운 경우 역행렬 계산을 크기 n인 행렬 대신 크기 k인 행렬에 대해 해도 M^{-1}을 구할 수 있음.

- A가 시간에 따른 함수인 경우 $dA^{-1}/dt = -A^{-1}(dA/dt)A^{-1}$임. A의 k번째 고유값 λ의 변화율은, A의 k번째 고유벡터 $\mathbf{x}$와 A'의 k번째 고유벡터 $\mathbf{y}$에 대한 식 $d\lambda/dt = \mathbf{y}'(dA/dt)\mathbf{x}$ 임. 특이값 σ의 변화율은 $d\sigma/dt = \mathbf{u}'(dA/dt)\mathbf{v}$ 임.