Texas A&M University

Landon Buechner

Prove that $\Omega_{i,j} = 0 \iff \beta_{ij} = 0$.

Let $\underline{Y}^T = (X_1, \ldots, X_d)^T$ be a random vector such that $X_i \sim \mathcal{N}(0, \sigma_i)$ for $i \in \{1, \ldots, d\}$. Assume that we can partition this into a vector $\underline{Z}$ whose entries are normal sub-vectors of $\underline{Y}$ such that $X_i \in \mathbb{R}$ and $\underline{Y}_i \in \mathbb{R}^{d-1}$.

$$Z = \begin{bmatrix} X_i \\ \underline{Y}_i \end{bmatrix} \sim \mathcal{N}(\vec{0},\ \Sigma) \tag{1}$$

Note that the covariance matrix can be block partitioned into a 2x2 matrix so that it contains the covariance information for the bivariate vector $\underline{Z}$. Assume that Σ_i and σ_i are invertible. The chosen partitioning implies that $\sigma_i \in \mathbb{R}$, $\vec{\sigma}_i \in \mathbb{R}^{d-1}$, and $\Sigma_i \in \mathbb{R}^{(d-1)^2}$. Since σ_i is a scalar this just means it must be a non-zero. In other words, X_i must be a non-degenerate distribution.

$$\Sigma = \begin{bmatrix} \sigma_i & \vec{\sigma}_i^T \\ \vec{\sigma}_i & \Sigma_i \end{bmatrix} \tag{2}$$

Recall that a bivariate joint distribution can be factorized as $f(x, y) = f(x|y)f(y)$. In the case of $\underline{Z}$, the form of the joint distribution is known.

$$f(\underline{Z}) = (2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}})^{-1}\exp\left\{ -\frac{1}{2}\underline{Z}^T\Sigma^{-1}\underline{Z} \right\} \tag{3}$$

It is unclear what Σ^{-1} is. With the assumptions about the entries in Σ stated above, the inverse of Σ can readily be found by diagonalizing Σ. Row reduction can be expressed as matrix operations such that matrices A and B left and right multiply Σ and result in the block diagonal matrix D, $A\Sigma B = D$. To find Σ^{-1}, invert both sides of the equation and solve.

$$\begin{aligned} (A\Sigma B)^{-1} &= D^{-1} \\ B^{-1}\Sigma^{-1}A^{-1} &= D^{-1} \\ \Sigma^{-1} &= BD^{-1}A \end{aligned} \tag{4}$$

Let I_k be a kxk identity matrix and see that $\mathrm{I}_1 = 1$. Additionally, define $\vec{0}_{pxq}$ as a pxq zero vector.

$$\begin{bmatrix} \mathrm{I}_1 & -\vec{\sigma}_i^T\Sigma_i^{-1} \\ \vec{0}_{(d-1)\text{x}1} & \mathrm{I}_{(d-1)} \end{bmatrix} \begin{bmatrix} \sigma_i & \vec{\sigma}_i^T \\ \vec{\sigma}_i & \Sigma_i \end{bmatrix} \begin{bmatrix} \mathrm{I}_1 & \vec{0}_{1\text{x}(d-1)} \\ -\Sigma_i^{-1}\vec{\sigma}_i & \mathrm{I}_{(d-1)} \end{bmatrix} = \begin{bmatrix} \sigma_i - \vec{\sigma}_i^T\Sigma_i^{-1}\vec{\sigma}_i & \vec{0}_{1\text{x}(d-1)} \\ \vec{0}_{(d-1)\text{x}1} & \Sigma_i \end{bmatrix} \tag{5}$$

Above, the specific dimensions are stated using the explicit notation and then are dropped in subsequent derivations because the dimensions can be easily inferred. This is further simplified by letting $\Sigma^* = \sigma_i - \vec{\sigma}_i^T\Sigma^{-1}\vec{\sigma}_i$.

$$\begin{aligned} \Omega = \begin{bmatrix} \sigma_i & \vec{\sigma}_i^T \\ \vec{\sigma}_i & \Sigma_i \end{bmatrix}^{-1} &= \begin{bmatrix} 1 & \vec{0} \\ -\Sigma_i^{-1}\vec{\sigma}_i & \mathrm{I} \end{bmatrix} \begin{bmatrix} (\Sigma^*)^{-1} & \vec{0} \\ \vec{0} & \Sigma_i^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\vec{\sigma}_i^T\Sigma_i^{-1} \\ \vec{0} & \mathrm{I} \end{bmatrix} \\ &= \begin{bmatrix} (\Sigma^*)^{-1} & -(\Sigma^*)^{-1}\vec{\sigma}_i^T\Sigma_i^{-1} \\ -\Sigma_i^{-1}\vec{\sigma}_i(\Sigma^*)^{-1} & \Sigma_i^{-1} + \Sigma_i^{-1}\vec{\sigma}_i(\Sigma^*)^{-1}\vec{\sigma}_i^T\Sigma_i^{-1} \end{bmatrix} \end{aligned} \tag{6}$$

Note that the block triangular matrices have determinants of one and recall that the determinant of a diagonal matrix is the product of the diagonal entries. Thus, $|\Sigma| = |\Sigma^*||\Sigma_i|$. This result is useful for expanding the normalizing factor in the joint density of $\underline{Z}$.

Letting $K = (2\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}})^{-1}$ and substituting the found inverse covariance matrix into the density it is simple to factor the joint distribution. An important property used in the below derivation is that for block partitioned matrices, $\begin{pmatrix} A \\ B \end{pmatrix}^T = \begin{pmatrix} A^T & B^T \end{pmatrix}$.

$$
\begin{aligned}
f(\underline{Z}) &= K\exp\Bigg\{ -\frac{1}{2}\begin{pmatrix} X_i \\ \underline{Y}_i \end{pmatrix}^T \begin{bmatrix} \sigma_i & \vec{\sigma}_i^T \\ \vec{\sigma}_i & \Sigma_i \end{bmatrix}^{-1} \begin{pmatrix} X_i \\ \underline{Y}_i \end{pmatrix} \Bigg\} \\
&= K\exp\Bigg\{ -\frac{1}{2}\begin{pmatrix} X_i^T & \underline{Y}_i^T \end{pmatrix} \begin{bmatrix} 1 & \vec{0} \\ -\Sigma_i^{-1}\vec{\sigma}_i & \mathrm{I} \end{bmatrix} \begin{bmatrix} (\Sigma^*)^{-1} & \vec{0} \\ \vec{0} & \Sigma_i^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\vec{\sigma}_i^T\Sigma_i^{-1} \\ \vec{0} & \mathrm{I} \end{bmatrix} \begin{pmatrix} X_i \\ \underline{Y}_i \end{pmatrix} \Bigg\} \\
&= K\exp\Bigg\{ -\frac{1}{2}\begin{pmatrix} X_i^T - \underline{Y}_i^T\Sigma_i^{-1}\vec{\sigma}_i & \underline{Y}_i^T \end{pmatrix} \begin{bmatrix} (\Sigma^*)^{-1} & \vec{0} \\ \vec{0} & \Sigma_i^{-1} \end{bmatrix} \begin{pmatrix} X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i \\ \underline{Y}_i \end{pmatrix} \Bigg\} \\
&= K\exp\Bigg\{ -\frac{1}{2}(X_i^T - \underline{Y}_i^T\Sigma_i^{-1}\vec{\sigma}_i)(\Sigma^*)^{-1}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i) - \frac{1}{2}\underline{Y}_i^T\Sigma_i^{-1}\underline{Y}_i \Bigg\} \\
&= K\exp\Bigg\{ -\frac{1}{2}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i)^T(\Sigma^*)^{-1}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i) - \frac{1}{2}\underline{Y}_i^T\Sigma_i^{-1}\underline{Y}_i \Bigg\} \\
&= K\exp\Bigg\{ -\frac{1}{2}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i)^T(\Sigma^*)^{-1}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i) \Bigg\} * \exp\Bigg\{ -\frac{1}{2}\underline{Y}_i^T\Sigma_i^{-1}\underline{Y}_i \Bigg\}
\end{aligned} \tag{7}
$$

Next see that K can be expanded into the exact normalizing factors for the conditional distribution $f(X_i|\underline{Y}_i)$ and the marginal $f(\underline{Y}_i)$ so that $f(\underline{Z}) = f(X_i|\underline{Y}_i)f(\underline{Y}_i)$.

$$
\begin{aligned}
f(X_i|\underline{Y}_i) &= (2\pi^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}})^{-1}\exp\Bigg\{ -\frac{1}{2}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i)^T(\Sigma^*)^{-1}(X_i - \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i) \Bigg\} \\
f(\underline{Y}_i) &= (2\pi^{\frac{d-1}{2}}|\Sigma|^{\frac{1}{2}})^{-1}\exp\Bigg\{ -\frac{1}{2}\underline{Y}_i^T\Sigma_i^{-1}\underline{Y}_i \Bigg\}
\end{aligned} \tag{8}
$$

Above shows how to find the conditional distributions for a normal random vector. In doing so, we have found the mean and variance. From here, it is natural to express X_i in terms of $\underline{Y}_i$ as a regression function.

$$
X_i|\underline{Y}_i \;\sim\; \mathcal{N}(\vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i,\ \Sigma^*) \tag{9}
$$

Recall that the covariance matrix is partitioned so that $\vec{\sigma}_i^T$ is a row vector of covariances between X_i and the remaining $d-1$ random variables. In other words. $\vec{\sigma}_i^T \equiv \mathrm{Cov}(X_i, \underline{Y}_i)$. Additionally, Σ_i^{-1} is the inverse covariance matrix for the $d-1$ dimension random vector $\underline{Y}_i$. Lastly note that that $\Sigma^* = \sigma_i - \vec{\sigma}_i^T\Sigma^{-1}\vec{\sigma}_i$. From here it is easy to see that X_i can be written as an equation that is linear in the coefficients that scale the elements of $\underline{Y}_i$. Let $\underline{\beta} \in \mathbb{R}^{1\mathrm{x}d-1}$ be such that $\beta_i = \vec{\sigma}_i^T\Sigma_i^{-1}$ and $\epsilon \sim \mathcal{N}(0, \sigma_i - \vec{\sigma}_i^T\Sigma^{-1}\vec{\sigma}_i)$.

$$
\begin{aligned}
X_i &= \vec{\sigma}_i^T\Sigma_i^{-1}\underline{Y}_i + \epsilon \\
&= \beta_i\underline{Y}_i + \epsilon \quad ,
\end{aligned} \tag{10}
$$

The work shown up to this point has been ultimately to show that $\Omega_{i,j} = 0 \iff \beta_{ij} = 0$. This is an important result when estimating the node wise neighborhoods in a Gaussian graphical model. Recall that $\beta_{i,j}$ is the k-th coefficient in β_i for equations of the same form as equation 10. Equivalently stated, the goal is to show that $\vec{\omega_{ik}}^T = 0 \iff \beta_{ik} = 0$. Here $\vec{\omega_{ik}}^T$ is the k-th element of the row vector of Ω characterized by the chosen partition scheme mentioned above. Below Ω is restated.

$$
\Omega = \begin{bmatrix} (\Sigma^*)^{-1} & -(\Sigma^*)^{-1}\vec{\sigma}_i^T\Sigma_i^{-1} \\ -\Sigma_i^{-1}\vec{\sigma}_i(\Sigma^*)^{-1} & \Sigma_i^{-1} + \Sigma_i^{-1}\vec{\sigma}_i(\Sigma^*)^{-1}\vec{\sigma}_i^T\Sigma_i^{-1} \end{bmatrix} = \begin{bmatrix} \omega_i & \vec{\omega}_i^T \\ \vec{\omega}_i & \Omega_i \end{bmatrix} \tag{11}
$$

Since we already know that $(\Sigma^*)^{-1} = \omega_i$ is non zero, notice the following equivalence.

$$\begin{aligned} \vec{\omega}_i^T &= -(\Sigma^*)^{-1} \vec{\sigma}_i^T \Sigma_i^{-1} = -(\Sigma^*)^{-1} \beta_i \\ &\Rightarrow \beta_i = -\frac{\vec{\omega}_i^T}{\omega_i} \end{aligned} \tag{12}$$

Since ω_i is non-zero, then when $\vec{\omega}_{ik} = 0$ then $\beta_{ik} = 0$ as well.