Normal–Inverse–Wishart Distribution

Juan F. Rubio-Ramírez

Emory University

Inverse Wishart and Matrix Gaussian Distributions

For $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$, positive definite matrices. Parameters: degrees of freedom $\nu > n - 1$, and scale matrix $\mathbf{\Psi} > 0$.

$$f_{IW}(\mathbf{\Sigma}; \nu, \mathbf{\Psi}) \propto |\, \mathbf{\Sigma} \,|^{-\frac{\nu+n+1}{2}} \exp\left(-\tfrac{1}{2}\mathrm{tr}(\mathbf{\Psi}\, \mathbf{\Sigma}^{-1})\right),$$

For $\mathbf{B} \in \mathbb{R}^{m \times n}$, if $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ are p.d. matrices of dimensions n and m respectively,

$$f_{MG}(\mathbf{B}; \boldsymbol{\mu}, \mathbf{\Sigma} \otimes \mathbf{\Omega}) \propto |\, \mathbf{\Sigma} \,|^{-\frac{m}{2}} \exp\left[-\tfrac{1}{2}\mathrm{tr}\left((\mathbf{B} - \boldsymbol{\mu})'\, \mathbf{\Omega}^{-1}(\mathbf{B} - \boldsymbol{\mu})\, \mathbf{\Sigma}^{-1}\right)\right].$$

Vector Autoregression (VAR) Setup

Let with

$$\mathbf{x}_t' = [\mathbf{y}_{t-1}' \ldots \mathbf{y}_{t-p}' \ \ 1] \text{ and } \mathbf{u}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}).$$

Reduced form:

$$\mathbf{y}_t' = \mathbf{x}_t' \, \mathbf{B} + \mathbf{u}_t'$$

Reduced-Form Likelihood

$$p(\mathbf{Y} \mid \mathbf{B}, \mathbf{\Sigma}) \propto |\,\mathbf{\Sigma}\,|^{-\frac{T}{2}} \exp\left[-\frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t' - \mathbf{x}_t' \mathbf{B}) \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{B}' \mathbf{x}_t) \right].$$

Using

$$\mathbf{Y}' = [\mathbf{y}_1 \ldots \mathbf{y}_T], \quad \mathbf{X}' = [\mathbf{x}_1 \ldots \mathbf{x}_T],$$

we get

$$p(\mathbf{Y} \mid \mathbf{B}, \mathbf{\Sigma}) \propto |\,\mathbf{\Sigma}\,|^{-\frac{T}{2}} \exp\Big[-\frac{1}{2} \mathrm{tr}\big((\mathbf{Y} - \mathbf{X}\,\mathbf{B})' (\mathbf{Y} - \mathbf{X}\,\mathbf{B}) \mathbf{\Sigma}^{-1} \big) \Big].$$

Reduced-Form Likelihood

The likelihood of the data given parameters $(\mathbf{B}, \mathbf{\Sigma})$ is:

$$p(\mathbf{Y} \mid \mathbf{B}, \mathbf{\Sigma}) = \prod_{t=1}^{T} p(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{B}, \mathbf{\Sigma}).$$

Since $\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{B}, \mathbf{\Sigma} \sim \mathcal{N}(\mathbf{B}' \mathbf{x}_t, \mathbf{\Sigma})$,

$$p(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{B}, \mathbf{\Sigma}) \propto= |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{y}_t' - \mathbf{x}_t' \mathbf{B}) \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{B}' \mathbf{x}_t)\right].$$

Combining terms over $t = 1, \ldots, T$:

$$p(\mathbf{Y} \mid \mathbf{B}, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-\frac{T}{2}} \exp\left[-\frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t' - \mathbf{x}_t' \mathbf{B}) \mathbf{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{B}' \mathbf{x}_t)\right].$$

Algebra: Expanding the Likelihood Trace Term

We start from the log-likelihood quadratic form:

$$\sum_{t=1}^{T}(\mathbf{y}_t' - \mathbf{x}_t' \mathbf{B}) \mathbf{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{B}' \mathbf{x}_t) = \mathrm{tr}(\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{B}' \mathbf{x}_t)(\mathbf{y}_t' - \mathbf{x}_t' \mathbf{B}) \mathbf{\Sigma}^{-1})$$

where we have used $\mathrm{tr}(\mathbf{X}\,\mathbf{Y}) = \mathrm{tr}(\mathbf{Y}\,\mathbf{X})$, then

$$\sum_{t=1}^{T}(\mathbf{y}_t' - \mathbf{x}_t' \mathbf{B}) \mathbf{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{B}' \mathbf{x}_t) = \mathrm{tr}((\mathbf{Y} - \mathbf{X}\,\mathbf{B})'(\mathbf{Y} - \mathbf{X}\,\mathbf{B}) \mathbf{\Sigma}^{-1})$$

where we have used $\mathbf{X}'\,\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i\,\mathbf{x}_i'$ and $\mathbf{X}' = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$.

Likelihood Simplification

Rewrite in matrix form:

$$\mathrm{tr}\left[(\mathbf{Y}-\mathbf{X}\,\mathbf{B})'(\mathbf{Y}-\mathbf{X}\,\mathbf{B})\,\mathbf{\Sigma}^{-1}\right] = \mathrm{tr}\left[\left(\hat{\mathbf{\Psi}} + (\mathbf{B}-\hat{\mu})'\hat{\mathbf{\Omega}}^{-1}(\mathbf{B}-\hat{\mu})\right)\mathbf{\Sigma}^{-1}\right] =$$

$$\mathrm{tr}\left[\hat{\mathbf{\Psi}}\,\mathbf{\Sigma}^{-1}\right] + \mathrm{tr}\left[(\mathbf{B}-\hat{\mu})'\hat{\mathbf{\Omega}}^{-1}(\mathbf{B}-\hat{\mu})\,\mathbf{\Sigma}^{-1}\right]$$

Define:

$$\hat{\mathbf{\Omega}} = (\mathbf{X}'\,\mathbf{X})^{-1}, \qquad \hat{\mu} = \hat{\mathbf{\Omega}}\,\mathbf{X}'\,\mathbf{Y}, \qquad \hat{\mathbf{\Psi}} = \mathbf{Y}'\,\mathbf{Y} - \hat{\mu}'\hat{\mathbf{\Omega}}^{-1}\hat{\mu}.$$

Then:

$$p(\mathbf{Y}\,|\,\mathbf{B},\mathbf{\Sigma}) \propto |\,\mathbf{\Sigma}\,|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(\hat{\mathbf{\Psi}}\,\mathbf{\Sigma}^{-1}\right) - \frac{1}{2}\mathrm{tr}\left((\mathbf{B}-\hat{\mu})'\hat{\mathbf{\Omega}}^{-1}(\mathbf{B}-\hat{\mu})\,\mathbf{\Sigma}^{-1}\right)\right].$$

Degrees of freedom. We need $|\,\mathbf{\Sigma}\,|^{-\frac{m}{2}}$ for the normal, hence we have $|\,\mathbf{\Sigma}\,|^{-\frac{T-m}{2}}|\,\mathbf{\Sigma}\,|^{-\frac{m}{2}}$.
This means $\hat{\upsilon} + n + 1 = T - m$, hence $\hat{\upsilon} = T - m - n - 1$.

Gaussian Inverse Wishart Prior

A likelihood:

$$p(\mathbf{Y} \mid \mathbf{B}, \boldsymbol{\Sigma}) = f_{IW}(\boldsymbol{\Sigma}; \hat{\nu}, \hat{\boldsymbol{\Psi}})\, f_{MG}(\mathbf{B}; \hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma} \otimes \hat{\boldsymbol{\Omega}}).$$

A conjugate prior:

$$p(\mathbf{B}, \boldsymbol{\Sigma}) = f_{IW}(\boldsymbol{\Sigma}; \bar{\nu}, \bar{\boldsymbol{\Psi}})\, f_{MG}(\mathbf{B}; \bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma} \otimes \bar{\boldsymbol{\Omega}}).$$

Implied posterior:

$$p(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}) = f_{IW}(\boldsymbol{\Sigma}; \tilde{\nu}, \tilde{\boldsymbol{\Psi}})\, f_{MG}(\mathbf{B}; \tilde{\boldsymbol{\mu}}, \boldsymbol{\Sigma} \otimes \tilde{\boldsymbol{\Omega}}),$$

where

$$\tilde{\nu} = T + \bar{\nu}, \quad \tilde{\boldsymbol{\Omega}} = (\hat{\boldsymbol{\Omega}}^{-1} + \bar{\boldsymbol{\Omega}}^{-1})^{-1}, \quad \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Omega}}(\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\mu}} + \bar{\boldsymbol{\Omega}}^{-1}\bar{\boldsymbol{\mu}}),$$
$$\tilde{\boldsymbol{\Psi}} = \hat{\boldsymbol{\Psi}} + \hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\mu}} + \bar{\boldsymbol{\Psi}} + \bar{\boldsymbol{\mu}}'\bar{\boldsymbol{\Omega}}^{-1}\bar{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}'\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\boldsymbol{\mu}}.$$

Completing the Square in the Matrix Case

Let $\mathbf{\Lambda}_i = \mathbf{\Omega}_i^{-1}$ (symmetric p.d.). Then

$$(\mathbf{B}-\mu_1)'\mathbf{\Lambda}_1(\mathbf{B}-\mu_1)+(\mathbf{B}-\mu_2)'\mathbf{\Lambda}_2(\mathbf{B}-\mu_2)=$$

$$\mathbf{B}'(\mathbf{\Lambda}_1+\mathbf{\Lambda}_2)\,\mathbf{B}-2\,\mathbf{B}'(\mathbf{\Lambda}_1\,\mu_1+\mathbf{\Lambda}_2\,\mu_2)+\mu_1'\,\mathbf{\Lambda}_1\,\mu_1+\mu_2'\,\mathbf{\Lambda}_2\,\mu_2$$

Choose $\mathbf{\Lambda}^* = \mathbf{\Lambda}_1+\mathbf{\Lambda}_2$ and $\mu^* = (\mathbf{\Lambda}^*)^{-1}(\mathbf{\Lambda}_1\,\mu_1+\mathbf{\Lambda}_2\,\mu_2)$. Then

$$(\mathbf{B}-\mu^*)'\mathbf{\Lambda}^*(\mathbf{B}-\mu^*)=\mathbf{B}'\,\mathbf{\Lambda}^*\,\mathbf{B}-2\,\mathbf{B}'\,\mathbf{\Lambda}^*\,\mu^*+(\mu^*)'\mathbf{\Lambda}^*\,\mu^*,$$

so $(\mathbf{B}-\mu_1)'\mathbf{\Lambda}_1(\mathbf{B}-\mu_1)+(\mathbf{B}-\mu_2)'\mathbf{\Lambda}_2(\mathbf{B}-\mu_2)=(\mathbf{B}-\mu^*)'\mathbf{\Lambda}^*(\mathbf{B}-\mu^*)+\mathbf{C}$, with $\mathbf{C}=\mu_1'\,\mathbf{\Lambda}_1\,\mu_1+\mu_2'\,\mathbf{\Lambda}_2\,\mu_2-(\mu^*)'\mathbf{\Lambda}^*\,\mu^*$.

The Degrees of Freedom

Finally, collect the powers of $|\boldsymbol{\Sigma}|$. The product has

$$-\frac{\nu_1+n+1}{2}-\frac{\nu_2+n+1}{2}-\frac{m}{2}-\frac{m}{2}=-\frac{\nu_1+\nu_2+(n+1)+m+(n+1)+m}{2}.$$

To express the result as one IW part and one matrix-normal part (i.e., one $|\boldsymbol{\Sigma}|^{-m/2}$ kept with the matrix-normal), set

$$-\frac{\nu^\star+n+1}{2}-\frac{m}{2}=-\frac{\nu_1+\nu_2+n+1+m+n+1}{2}-\frac{m}{2},$$

which gives the updated degrees of freedom $\nu^\star=\nu_1+\nu_2+n+1+m$. In our case $\nu_1=\bar{\nu}$, $\nu_2=\hat{\nu}=T-m-n-1$, and $\nu^*=\tilde{\nu}$, hence $\tilde{\nu}=T+\bar{\nu}$.