ADVANCED BAYESIAN MODELING

Inverse Wishart Prior

In a varying coefficient model, coefficient vectors $\beta^{(j)}$ may have arbitrary (positive definite) covariance matrix:

$$\mathrm{var}\big(\beta^{(j)} \mid \mu_\beta, \Sigma_\beta\big) \;\; = \;\; \Sigma_\beta$$

Generally, data should be used to estimate Σ_β, so it needs a prior.

What would be a natural choice for the distribution of a positive definite matrix?

Wishart Distribution

A mathematically convenient distribution on positive definite matrices is the **Wishart distribution**

$$\mathrm{Wishart}_\nu(S)$$

with ν degrees of freedom, and positive definite scale matrix S.

If

$$W \sim \mathrm{Wishart}_\nu(S) \qquad \text{is } k \times k$$

then, for $\nu > k - 1$,

$$p(W) \propto |W|^{(\nu-k-1)/2}\, e^{-\operatorname{tr}(S^{-1}W)/2} \qquad W \text{ positive definite}$$

$$E(W) = \nu S$$

(There are Wishart distributions with $\nu \leq k-1$, but they are degenerate.)

Wishart generalizes the gamma distribution:

$$\mathrm{Gamma}(\alpha, \beta) \qquad = \qquad \mathrm{Wishart}_{2\alpha}\big((2\beta)^{-1}\big)$$

Recall: Scaled inverse chi-square is (partially) conjugate for a normal variance.

Since this is a type of inverse gamma, perhaps a (partially) conjugate prior for a multivariate normal covariance matrix might be *inverse* Wishart ...

Inverse Wishart Distribution

If $W \sim \text{Wishart}_\nu(S)$ then W^{-1} has the **inverse Wishart distribution**:

$$W^{-1} \sim \text{Inv-Wishart}_\nu(S)$$

(in parameterization of BDA3, Sec. A.1)

The inverse Wishart is conjugate for a normal covariance matrix.

Indeed, it generalizes the scaled inverse chi-square:

$$\text{Inv-}\chi^2(\nu, s^2) \qquad = \qquad \text{Inv-Wishart}_\nu\big((\nu s^2)^{-1}\big)$$

Inverse Wishart on $k \times k$ matrices exists only when $\nu > k - 1$.

As a prior, it is less informative when ν is smaller.

Typical default choices for ν:

- $\nu = k$: smallest possible integer value
- $\nu = k + 1$: recommended in BDA3 – correlations are marginally uniform

Remark: Setting ν almost to $k - 1$ is *not* recommended for a hyperprior.

Inverse Wishart Hyperprior

Could choose

$$\Sigma_\beta \sim \text{Inv-Wishart}_K\left(\Sigma_0^{-1}/K\right)$$

or

$$\Sigma_\beta \sim \text{Inv-Wishart}_{K+1}\left(\Sigma_0^{-1}/(K+1)\right)$$

for both of which Σ_β^{-1} has mean Σ_0^{-1}.

Criticized in BDA3 for being too constrained (informative).

In practice, results can be sensitive to choice of Σ_0 – it's tempting to choose Σ_0 based on the data.

In JAGS

JAGS favors using precision matrix Σ_β^{-1}:

$$\Sigma_\beta^{-1} \sim \text{Wishart}_K\left(\Sigma_0^{-1}/K\right)$$

or

$$\Sigma_\beta^{-1} \sim \text{Wishart}_{K+1}\left(\Sigma_0^{-1}/(K+1)\right)$$

The first might appear in a JAGS model as

```
Sigmabetainv ~ dwish(K * Sigma0, K)
```

Note:

- Ordering of arguments to `dwish`
- Matrix argument is $K\Sigma_0$ instead of Σ_0^{-1}/K

Scaled Inverse Wishart

BDA3 instead recommends using

$$\Sigma_\beta \;=\; \mathrm{Diag}(\xi)\,\Sigma_\eta\,\mathrm{Diag}(\xi)$$

$$\Sigma_\eta \;\sim\; \text{Inv-Wishart}_K(I) \qquad\qquad \xi \;\sim\; \text{diffuse or flat on } \mathbb{R}_+^K$$

Advantage: No need for sensitive choice of Σ_0.

Can be implemented as

$$\beta^{(j)} \;=\; \mu_\beta + \xi \circ \eta^{(j)} \qquad\qquad \eta^{(j)} \;\sim\; \mathrm{N}(0, \Sigma_\eta)$$

where $\circ$ is elementwise product.