

The Matrix-variate F prior

Matt Simpson

February 11, 2014

1 Mixing Wisharts and inverse Wisharts

The main goal of this section is to make precise and prove the statement “an inverse Wishart mixture of Wisharts is the same as a Wishart mixture of inverse Wisharts” as well as describe some of the properties of the resulting class of distributions. We will work with a slightly more general class of distribution in order to somewhat simplify the mathematics – the (inverse) matrix gamma distribution. Let X be a $p \times p$ nonnegative definite random matrix with a matrix gamma distribution given shape and scale parameters $\alpha > (p-1)/2$ and $\beta > 0$, and scale matrix parameter Σ , a $p \times p$ positive definite matrix, i.e. $X \sim MG_p(\alpha, \beta, \Sigma)$. Then the density of X is

$$p(X) = \frac{|\Sigma|^{-\alpha}}{\beta^{p\alpha} \Gamma_p(\alpha)} |X|^{\alpha-(p+1)/2} \exp \left[\text{tr} \left(-\frac{1}{\beta} \Sigma^{-1} X \right) \right]$$

Where $|\cdot|$ denotes the determinant operator, $\text{tr}(\cdot)$ the trace operator, and $\Gamma_p(\cdot)$ is the multivariate gamma function. $Y = X^{-1}$ has an inverse matrix gamma distribution, $Y \sim IMG_p(\alpha, \beta, \Psi)$ with $\Psi = \Sigma^{-1}$, and the density of Y is

$$p(Y) = \frac{|\Psi|^\alpha}{\beta^{p\alpha} \Gamma_p(\alpha)} |Y|^{-\alpha-(p+1)/2} \exp \left[\text{tr} \left(-\frac{1}{\beta} \Psi Y^{-1} \right) \right].$$

The (inverse) Wishart distribution is a special case with $\alpha = n/2$ and $\beta = 0$, where n is the degrees of freedom parameter of the (inverse) Wishart distribution, denoted $X \sim W_p(n, \Sigma)$ and $Y \sim IW_p(n, \Psi)$. Now we can state the two main theorems of this section.

Theorem 1.0.1. *Suppose $X|Y \sim MG_p(\alpha_1, \beta_1, Y)$ and $Y \sim IMG_p(\alpha_2, \alpha_2, \Sigma)$ Then the marginal distribution of X is*

$$p(X) = \frac{\left| \frac{\beta_1}{\beta_2} \Sigma \right|^{\alpha_2}}{B_p(\alpha_1, \alpha_2)} |X|^{\alpha_1-(p+1)/2} \left| X + \frac{\beta_1}{\beta_2} \Sigma \right|^{-(\alpha_1+\alpha_2)}$$

Theorem 1.0.2. *Suppose $X|Y \sim IMG_p(\alpha_2, \beta_2, Y)$ and $Y \sim MG_p(\alpha_1, \beta_1, \Sigma)$ Then the marginal distribution of X is*

$$p(X) = \frac{\left| \frac{\beta_1}{\beta_2} \Sigma \right|^{\alpha_2}}{B_p(\alpha_1, \alpha_2)} |X|^{\alpha_1-(p+1)/2} \left| X + \frac{\beta_1}{\beta_2} \Sigma \right|^{-(\alpha_1+\alpha_2)}$$

where $B_p(\alpha_1, \alpha_2)$ is the multivariate beta function with the property $B_p(\alpha_1, \alpha_2) = \Gamma_p(\alpha_1) \Gamma_p(\alpha_2) / \Gamma_p(\alpha_1 + \alpha_2)$.

The proofs of these theorems are actually quite simple. We'll start with Theorem (1.0.1). The joint distribution of (X, Y) can be written as

$$p(X, Y) = \frac{|\Sigma|^{\alpha_2} \Gamma_p(\alpha_1 + \alpha_2)}{(\beta_1^{\alpha_1} \beta_2^{\alpha_2})^p \Gamma_p(\alpha_1) \Gamma_p(\alpha_2)} \left| \frac{1}{\beta_1} X + \frac{1}{\beta_2} \Sigma \right|^{-(\alpha_1 + \alpha_2)} |X|^{\alpha_1 - (p+1)/2} \\ \times \frac{\left| \frac{1}{\beta_1} X + \frac{1}{\beta_2} \Sigma \right|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |Y|^{-(\alpha_1 + \alpha_2) - (p+1)/2} \exp \left[\text{tr} \left(- \left(\frac{1}{\beta_1} X + \frac{1}{\beta_2} \Sigma \right) Y^{-1} \right) \right].$$

The second line is the density of an inverse matrix gamma distribution, so marginalizing out Y and rearranging a bit we get

$$p(X) = \frac{\left| \frac{\beta_1}{\beta_2} \Sigma \right|^{\alpha_1}}{B_p(\alpha_1, \alpha_2)} |X|^{\alpha_1 - (p+1)/2} \left| X + \frac{\beta_1}{\beta_2} \Sigma \right|^{-(\alpha_1 + \alpha_2)}$$

which completes the proof.

To prove Theorem (1.0.2) we'll do something similar but ignore the normalizing constant. In this case the joint distribution of X and Y can be written as

$$P(X, Y) \propto |X|^{-\alpha_2 - (p+1)/2} |Y|^{\alpha_1 - (p+1)/2} \exp \left[\text{tr} \left(- \left(\frac{1}{\beta_2} X^{-1} + \frac{1}{\beta_1} \Sigma^{-1} \right) Y \right) \right].$$

Now we marginalize out Y to get the desired result:

$$p(X) \propto |X|^{-\alpha_2 - (p+1)/2} \left| \frac{1}{\beta_2} X^{-1} + \frac{1}{\beta_1} \Sigma^{-1} \right|^{-(\alpha_1 + \alpha_2)} \\ \propto |X|^{\alpha_1 - (p+1)/2} \left| \frac{\beta_1}{\beta_2} \Sigma + X \right|^{-(\alpha_1 + \alpha_2)}$$

where the last line comes from $|X \beta_1 \Sigma|^{\alpha_1 + \alpha_2} |X \beta_1 \Sigma|^{-(\alpha_1 + \alpha_2)} = 1$.

When $\alpha_1 = 2n_1$, $\alpha_2 = 2n_2$ and $\beta_1 = \beta_2$ (or both β 's are absorbed into Σ), X has the multivariate (or matrix-variate) F distribution ($X \sim F_p(n_1, n_2, \Sigma)$) with density

$$p(X) \propto |X|^{\frac{n_1 - (p+1)}{2}} |I_p + \Sigma^{-1} X|^{-(n_1 + n_2)/2}$$

where I_p is the $p \times p$ identity matrix. We'll focus on the multivariate F distribution as a prior for covariance matrices, specifically $X \sim F_p(n_1, n_2 + p - 1, \Sigma)$.

2 Properties of the multivariate F prior on covariance matrices

The multivariate F distribution is well known in the multivariate analysis literature. Dawid [1981] defines it as an inverse Wishart mixture of Wisharts, as in Theorem (1.0.1), and notes that it can be seen as a generalization of the Wishart distribution. The following corollary illustrates this.

Corollary 2.0.3. *Suppose for $i = 1, 2, \dots, n_1$, x_i is a p -dimensional random vector with*

$$x_i \stackrel{iid}{\sim} T_p(n_2 - p + 1, 0, \Sigma / (n_2 - p + 1))$$

where $T_p(\nu, \mu, \Omega)$ is the multivariate T distribution with degrees of freedom ν , $p \times 1$ location parameter μ and $p \times p$ scale matrix Ω . Suppose $n_1, n_2 > p - 1$ and let $S = \sum_{i=1}^{n_1} x_i x_i'$. Then $S \sim F_p(n_1, n_2, \Sigma)$

To see this, recall that if $x|Y \sim N_p(0, Y)$ and $Y \sim IW_p(\nu, \Sigma)$ then marginally $x \sim T_p(\nu - p + 1, 0, \Sigma/(\nu - p + 1))$. So we can write the distribution of the x_i 's as independent inverse Wishart mixtures of independent normals, i.e. $x_i|Y_1, Y_2, \dots, Y_{n_1} \stackrel{ind}{\sim} N_p(0, Y_i)$ and $Y_i \stackrel{iid}{\sim} IW_p(n_2, \Sigma)$. But conditional on Y , we know $S \sim W_p(n_1, Y)$. So by Theorem (1.0.1) $X \sim F_p(n_1, n_2, \Sigma)$.

This family of covariance matrix priors has the same self-consistent property that Huang and Wand [2013] discuss.

Proposition 2.0.4. *If $X \sim F_p(n_1, n_2 + p - 1, \Sigma)$ with*

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{12} & X_{22} \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}$$

where X_{11} is a $q \times q$ submatrix of X and Σ_{11} is a $q \times q$ submatrix of Σ . Then $X_{11} \sim F_q(n_1, n_2 + q - 1, \Sigma_{11})$

Using Theorem (1.0.1) we have $X|Y \sim W_p(n_1, Y)$ and $Y \sim IW_p(n_2 + p - 1, \Sigma)$. Recall that for Wishart random matrices that if C is a $q \times p$ rank q matrix then $CXC' \sim W_q(n_1, CYC')$. Set $C = (c'_1, c'_2, \dots, c'_q)'$ with c_q a row vector of zeroes except for a one in the q 'th place. Then we have $X_{11}|Y_{11} \sim W_q(n_1, Y_{11})$. Now from the properties of the inverse Wishart distribution $Y_{11} \sim IW_q(n_2 + p - 1 - (p - q), \Sigma_{11})$. Then using Theorem (1.0.1) this gives $X_{11} \sim F_q(n_1, n_2 + p - 1, \Sigma_{11})$.

This self consistency property is desirable because it allows us to expand a covariance matrix without changing the marginal distribution of the original matrix. As Huang and Wand [2013] note, this property is nontrivial and does not hold for some other proposed covariance matrix priors, e.g. Barnard et al. [2000]. This property does hold for the inverse Wishart prior, as well as another nice property — if we prefer to work with the precision matrix instead of the covariance matrix, it's easy to do so. The next proposition establishes this fact for the proposed family of priors.

The previous set of properties of the matrix-variate F distribution can be found in Dawid [1981]. The next result may be novel though.

Proposition 2.0.5. *Suppose $X \sim F_p(n_1 + p - 1, n_2 + p - 1, \Sigma)$. Then $X^{-1} \sim F_p(n_2 + p - 1, n_1 + p - 1, \Sigma^{-1})$*

Using Theorem (1.0.1) we have $X|Y \sim W_p(n_1 + p - 1, Y)$ and $Y \sim IW_p(n_2 + p - 1, \Sigma)$. Then $X^{-1}|Y \sim IW_p(n_1 + p - 1, Y^{-1})$ and $Y^{-1} \sim W_p(n_2 + p - 1, \Sigma^{-1})$. So using Theorem (1.0.2) $X^{-1} \sim F_p(n_2 + p - 1, n_1 + p - 1, \Sigma)$. Thus we have a choice — we can make this family of priors self consistent on the covariances or self consistent on the precisions, but not both. We are interested in priors for covariance matrices so we enforce the self consistency property on that scale.

We'll consider the marginal distribution of the standard deviation parameters with the next proposition.

Proposition 2.0.6. *If $X \sim F_p(n_1, n_2 + p - 1, \Sigma)$ then the marginal distribution of $s_{ii} = \sqrt{x_{ii}}$ is the “scale mixed Nakagami” distribution with density*

$$p(s_{ii}) \propto s_{ii}^{n_1-1} \left(1 + \frac{s_{ii}^2}{\sigma_{ii}} \right)^{-(n_1+n_2)/2}.$$

Using the self consistency property we immediately have

$$p(x_{ii}) \propto x_{ii}^{n_1/2-1} \left(1 + \frac{x_{ii}}{\sigma_{ii}} \right)^{-(n_1+n_2)/2}.$$

Now let $x_{ii} = s_{ii}^2$ with Jacobian $2s_{ii}$ we get

$$p(s_{ii}) \propto s_{ii}^{n_1-1} \left(1 + \frac{s_{ii}^2}{\sigma_{ii}} \right)^{-(n_1+n_2)/2}.$$

Note that since $n_1 > p - 1$, we cannot in general have the marginal distribution of the standard deviations be half- t . We are calling this a “scale mixed Nakagmi” because the Nakagami distribution, under one parameterization, is

$$p(x) = \frac{1}{2^{\alpha-1}\theta^\alpha\Gamma(\alpha)}x^{2\alpha-1}\exp\left[-\frac{x^2}{2\theta}\right].$$

By choosing $\alpha = n_1/2$ and setting $\theta \sim IG(n_2/2, \sigma_{ii}/2)$ with pdf $p(\theta) \propto \theta^{-n_2/2-1}\exp(2\sigma_{ii}/\theta)$, then marginalizing out θ , we obtain the marginal distribution of the standard deviations of the F_p distribution.

Characterizing the marginal distribution of the correlation parameters has proven to be more difficult. However, it’s not difficult to simulate from this distribution and use the results to choose n_1 and n_2 so that the correlations are marginally uniform, supposing that Σ is diagonal.

LOOK AT WHAT HAPPENS WHEN EITHER n_1 OR n_2 GO TO ∞ . WHEN $n_2 \rightarrow \infty$ WE GET THE WISHART. SUSPECT WHEN $n_2 \rightarrow \infty$ WE GET THE INVERSE WISHART

EXPECTED VALUE AND VARIANCE OF SD, VARIANCE, AND/OR CORRELATION PARAMETERS

DEPENDENCE BETWEEN SD/VARIANCE AND CORRELATION - EARLY SIMULATIONS SHOW LITTLE DEPENDENCE BUT STILL A BIT - A BIT OF POSITIVE CORRELATION BETWEEN ρ^2 and σ^2

3 Discussion

The significance of this distribution and its mixture representations is hopefully clear — the matrixvariate F distribution provides a much more general class of prior distributions for both covariance and precision matrices than typically used while still allowing computation to be cheap and easily implemented. The main cost here is an additional Gibbs step where we draw from an (inverse) Wishart distribution for the scale matrix in the prior for the covariance (precision) matrix. A similar idea has been put forth by Huang and Wand [2013] with a similar cost. Their suggestion is to use an inverse Wishart prior for the covariance matrix, but force scale matrix parameter to be diagonal. Then put independent gamma hyperpriors on the diagonal elements of the scale matrix. They note that this allows for a more flexible covariance matrix prior, including marginal half- t distributions for standard deviations and, if desired, marginal uniform distributions for the correlations. One advantage of our prior, like Huang and Wand’s, is that it is easy to construct an MCMC sampler using the precision matrix instead the covariance matrix due to Theorem (1.0.1).

References

- John Barnard, Robert McCulloch, and Xiao-Li Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1312, 2000.
- A Philip Dawid. Some matrix-variate distribution theory: notational considerations and a bayesian application. *Biometrika*, 68(1):265–274, 1981.
- Alan Huang and MP Wand. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452, 2013.