

# The matrix F prior

Matt Simpson

February 17, 2014

## 1 The matrix F distribution

The matrix F distribution is a well known distribution in random matrix theory and in multivariate analysis where it is related to the matrix-variate analogue of ratios of sums of squares and can be seen as a matrix-variate analogue to the beta distribution. For example, see Dawid [1981], ?, and also ? where it is an example of the generalized type II matrix-variate beta distribution. For  $X \sim F_p(n_1, n_2, \Sigma)$  where  $n_1, n_2 \in \{1, 2, \dots, p-1\} \cup (p-1, \infty)$  are degrees of freedom parameters with  $\max\{n_1, n_2\} > p-1$ , and  $\Sigma$  a symmetric, positive definite scale matrix, the density of  $X$  is

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

The normalizing constant depends on whether or not  $n_1, n_2 > p-1$ . In section 3 we show that when  $n_1 = 1$  the marginal distribution of the standard deviation parameters in  $X$  is half- $t$ , which motivates why we're considering small  $n_1$ .

There are a number of different methods to derive the density of the matrix F distribution. We'll use one method in two different ways because of the relevance to selecting priors on covariance matrices. The basic idea is to take  $X|Y$  as a draw from a Wishart distribution with scale matrix  $Y$ , while  $Y$  as a draw from an inverse Wishart distribution. There are two ways to define the Wishart distribution. The first comes from the matrix normal distribution – suppose  $Z$  is an  $n \times p$  random matrix drawn from a zero mean matrix normal distribution with independent rows and dependent columns, i.e.  $Z \sim N_{n,p}(0_{n \times p}, I_n, \Sigma)$  where  $0_{n \times p}$  is an  $n \times p$  matrix of zeroes. Then  $X = Z'Z$  has a Wishart distribution with  $n$  degrees of freedom and symmetric positive definite scale matrix  $\Sigma$ , i.e.  $X \sim W_p(n, \Sigma)$ . When  $n > p-1$  it can be shown that  $X$  has the density

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

where  $\text{tr}(\cdot)$  is the trace operator and  $\Gamma_p$  is the multivariate gamma function. This function can be written in terms of univariate gamma functions:

$$\Gamma_p(\alpha) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(\alpha - (i-1)/2)$$

When  $n \leq p-1$  the distribution of  $X$  is singular with respect to Lebesgue measure. The second method of defining the Wishart distribution is identifying it with the above density, which can be extended to noninteger  $n > p-1$ .

The inverse Wishart distribution is defined as the distribution of the inverse of a Wishart distributed matrix. Suppose  $X \sim W_p(n, \Sigma)$  then  $Y = X^{-1}$  has an inverse Wishart distribution, i.e.  $Y \sim IW_p(n, \Psi)$  where  $\Psi = \Sigma^{-1}$ . If the underlying Wishart distribution has degrees of freedom  $n > p-1$  then  $Y$  has the density

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

Also, for reference, if  $X \sim N_{n,p}(M, U, V)$  then  $X$  has the density

$$p(X) = \frac{\exp \left[ -\frac{1}{2} \text{tr} [V^{-1}(X - M)'U^{-1}(X - M)] \right]}{(2\pi)^{np/2} |V|^{n/2} |U|^{p/2}}.$$

## 2 Mixing Wisharts and inverse Wisharts

The main goal of this section derive the matrix F distribution as an inverse Wishart mixture of Wisharts as well as a Wishart mixture of inverse Wisharts. We'll start with the case where both  $n_1, n_2 > p - 1$  with the following two theorems.

**Theorem 2.0.1.** *Suppose  $X|Y \sim W_p(n_1, Y)$  and  $Y \sim IW_p(n_2, \Sigma)$  Then the marginal distribution of  $X$  is*

$$p(X) = \frac{|\Sigma|^{n_2/2}}{B_p(n_1/2, n_2/2)} |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2}$$

**Theorem 2.0.2.** *Suppose  $X|Y \sim IW_p(n_2, Y)$  and  $Y \sim W_p(n_1, \Sigma)$  Then the marginal distribution of  $X$  is*

$$p(X) = \frac{|\Sigma|^{n_2/2}}{B_p(n_1/2, n_2/2)} |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2}$$

Here  $B_p(\alpha_1, \alpha_2)$  is the multivariate beta function with the property

$$B_p(\alpha_1, \alpha_2) = \frac{\Gamma_p(\alpha_1)\Gamma_p(\alpha_2)}{\Gamma_p(\alpha_1 + \alpha_2)}.$$

The proofs of these theorems are actually quite simple. We'll start with Theorem (2.0.1). The joint distribution of  $(X, Y)$  can be written as

$$\begin{aligned} p(X, Y) &= |\Sigma|^{n_2/2} \frac{\Gamma_p((n_1 + n_2)/2)}{\Gamma_p(n_1/2)\Gamma_p(n_2/2)} \frac{2^{(n_1+n_2)p/2}}{2^{n_1p/2}2^{n_2p/2}} |X + \Sigma|^{-(n_1+n_2)/2} |X|^{(n_1-p-1)/2} \\ &\times \frac{|X + \Sigma|^{(n_1+n_2)/2}}{\Gamma_p((n_1 + n_2)/2)2^{(n_1+n_2)p/2}} |Y|^{-(n_1+n_2-p-1)/2} \exp \left[ -\frac{1}{2} \text{tr} ((X + \Sigma)Y^{-1}) \right]. \end{aligned}$$

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

$$p(X) = \frac{|\Sigma|^{n_1/2}}{B_p(n_1/2, n_2/2)} |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2}$$

which completes the proof.

To prove Theorem (2.0.2) we'll do something similar but ignore the normalizing constant. In this case the joint distribution of  $X$  and  $Y$  can be written as

$$P(X, Y) \propto |X|^{-(n_2-p-1)/2} |Y|^{(n_1-p-1)/2} \exp \left[ -\frac{1}{2} \text{tr} ((X^{-1} + \Sigma^{-1})Y) \right].$$

Now we marginalize out  $Y$  to get the desired result:

$$\begin{aligned} p(X) &\propto |X|^{-(n_2-p-1)/2} |X^{-1} + \Sigma^{-1}|^{-(n_1+n_2)/2} \\ &\propto |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2} \end{aligned}$$

where the last line comes from  $|X\Sigma|^{(n_1+n_2)/2} |X\Sigma|^{-(n_1+n_2)/2} = 1$ .

The next two theorems allow for possibly singular densities in the mixing distribution, but not in the conditional distribution.

**Theorem 2.0.3.** Suppose  $X|Y \sim IW_p(n_2, Y'Y)$  and  $Y \sim N_{n_1,p}(0_{n_1 \times p}, I_{n_1}, \Sigma)$  potentially with  $n_1 < p$  but  $n_2 > p - 1$ . Then the marginal density of  $X$  is

$$p(X) = \frac{|\Sigma|^{n_2/2}}{\Gamma_p(n_2/2)2^{n_2p/2}} E[|Z'Z|^{n_2/2}] |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2}$$

where  $Z$  is an  $n_1 \times p$  matrix of iid standard normally distributed random variables.

**Theorem 2.0.4.** Suppose  $X|Y \sim W_p(n_1, (Y'Y)^{-1})$  and  $Y \sim N_{n_2,p}(0_{n_2 \times p}, I_{n_2}, \Sigma^{-1})$  potentially with  $n_2 < p$  but  $n_1 > p - 1$ . Then the marginal density of  $X$  is

$$p(X) = \frac{|\Sigma|^{n_2/2}}{\Gamma_p(n_1/2)2^{n_1p/2}} E[|Z'Z|^{n_1/2}] |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2}$$

where  $Z$  is an  $n_2 \times p$  matrix of iid standard normally distributed random variables.

Theorems (2.0.3) and (2.0.4) are a little bit harder to prove, but not much. We will only prove Theorem (2.0.3) since the proof of Theorem (2.0.3) is analogous. We can write the joint distribution of  $X$  and  $Y$  as

$$\begin{aligned} p(X, Y) &= \frac{|\Sigma|^{-n_1/2}}{2^{n_2p/2} \Gamma_p(n_2/2)} |X|^{-(n_2+p+1)/2} (2\pi)^{-n_1p/2} |Y'Y|^{n_2/2} \exp \left[ -\frac{1}{2} \text{tr}((\Sigma^{-1} + X^{-1})Y'Y) \right] \\ &= \frac{|\Sigma|^{-n_1/2}}{2^{n_2p/2} \Gamma_p(n_2/2)} |X|^{-(n_2+p+1)/2} |X^{-1} + \Sigma^{-1}|^{-n_1/2} \\ &\quad \times |X^{-1} + \Sigma^{-1}|^{n_1/2} (2\pi)^{-n_1p/2} |Y'Y|^{n_2/2} \exp \left[ -\frac{1}{2} \text{tr}((\Sigma^{-1} + X^{-1})Y'Y) \right]. \end{aligned}$$

The last line constains the kernel of a matrix normal distribution. This allows us to write the marginal distribution of  $X$  as

$$\begin{aligned} p(X) &= \frac{|\Sigma|^{-n_1/2}}{2^{n_2p/2} \Gamma_p(n_2/2)} |X|^{-(n_2+p+1)/2} |X^{-1} + \Sigma^{-1}|^{-n_1/2} E[|Y'Y|^{n_2/2} | X] \\ &= \frac{|\Sigma|^{-n_1/2}}{2^{n_2p/2} \Gamma_p(n_2/2)} |X|^{-(n_2+p+1)/2} |X^{-1} + \Sigma^{-1}|^{-n_1/2} E[|(X^{-1} + \Sigma^{-1})^{-1/2} Y'Y (X^{-1} + \Sigma^{-1})^{-1/2}|^{n_2/2} | X] \\ &= \frac{|\Sigma|^{-n_1/2} E[|Z'Z|^{n_2/2}]}{2^{n_2p/2} \Gamma_p(n_2/2)} |X|^{-(n_2+p+1)/2} |X^{-1} + \Sigma^{-1}|^{-(n_1+n_2)/2} \\ &= \frac{|\Sigma|^{n_2/2} E[|Z'Z|^{n_2/2}]}{2^{n_2p/2} \Gamma_p(n_2/2)} |X|^{(n_1-p-1)/2} |X + \Sigma|^{-(n_1+n_2)/2} \end{aligned}$$

where  $Z \sim N_{n_1,p}(0_{n_1 \times p}, I_{n_2}, I_p)$ . The second to last line comes from the fact that if  $Z \sim N_{n,p}(0_{n \times p}, I_n, I_p)$  then  $ZB^{1/2} \sim N_{n,p}(0_{n \times p}, I_n, B)$ . When  $n_1 > p - 1$ ,  $E[|Z'Z|^{n_2/2}]$  can be explicitly computed using the Bartlett decomposition and will agree with Theorems (2.0.2) and (2.0.1). When  $n_1 \leq p - 1$  the expectation and thus normalizing constant are finite since all moments exist for the normal distribution.

### 3 Properties of the multivariate F prior on covariance matrices

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

**Proposition 3.0.5.** 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  $\Sigma_{11}$  is a  $q \times q$  submatrix of  $\Sigma$ . Then  $X_{11} \sim F_q(n_1, n_2 + q - 1, \Sigma_{11})$

Using Theorems (2.0.2) and (2.0.3) we have  $X|Y \sim IW_p(n_2 + p - 1, Y)$  and  $Y \sim W_p(n_1, \Sigma)$ . If  $n_1 \leq p - 1$  then  $Y = Z'Z$  with  $Z \sim N_{n_1, p}(0_{n_1 \times p}, I_{n_1}, \Sigma)$ . From the properties of the inverse Wishart distribution  $X_{11}|Y_{11} \sim IW_q(n_2 + p - 1 - (p - q), Y_{11})$ . 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  $Y_{11} \sim W_q(n_1, \Sigma_{11})$ . Then using either Theorem (2.0.2) or (2.0.3) this gives  $X_{11} \sim F_q(n_1, n_2 + p - 1, \Sigma_{11})$ .

This property of the matrix F distribution can be seen in Dawid [1981] for  $n_1 > p - 1$ . The 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.

**Proposition 3.0.6.** *Suppose  $X \sim F_p(n_1, n_2, \Sigma)$  with  $\max\{n_1, n_2\} > p - 1$ . Then  $X^{-1} \sim F_p(n_2, n_1, \Sigma^{-1})$*

The Jacobian of the transformation  $X \rightarrow Y = X^{-1}$  is  $|Y|^{-(p+1)}$  (see ? for details on how to show this). Then

$$\begin{aligned} p(Y) &\propto |Y|^{-(p+1)} |Y^{-1}|^{-(n_1+p+1)/2} |\Sigma + Y^{-1}|^{-(n_1+n_2)/2} \\ &\propto |Y|^{(n_2-p-1)/2} |Y + \Sigma^{-1}|^{-(n_1+n_2)/2}. \end{aligned}$$

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

**Proposition 3.0.7.** *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

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

When we set  $n_1 = 1$ , then the diagonal elements of  $X$  have marginal half- $t$  distributions. 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  $\theta$ , 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  $\Sigma$  is diagonal.

*LOOK AT WHAT HAPPENS WHEN EITHER  $n_1$  OR  $n_2$  GO TO  $\infty$ . 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  $\rho^2$  and  $\sigma^2$

### 3.1 Implementation in a Gibbs sampler

Suppose we have some model with data  $y|\theta, \Sigma \sim p(y|\theta, \Sigma)$  where  $\theta$  is some unknown parameter and  $\Sigma$  is a  $p \times p$  covariance matrix. If we suppose that a prior  $\theta$  and  $\Sigma$  are independent and use the matrix F prior on  $\Sigma$ , then we can decompose this prior using Theorem (2.0.3) to write  $\Sigma|Z \sim IW_p(n_2 + p - 1, Z'Z)$  and  $Z \sim N_{n_1, p}(0_{n_1 \times p}, I_{n_1}, \Omega)$ . If  $\Sigma$  is the covariance matrix of a normal distribution, we have a nice conditionally conjugate Gibbs step for drawing  $\Sigma$  given the data and all other parameters, including  $Z$ , from an inverse Wishart distribution. Then the conditional distribution of  $Z$  given the data and all model parameters depends only on  $\Sigma$  and is

$$p(Z) \propto |Z'Z|^{n_2/2} \exp \left[ -\frac{1}{2} \text{tr}((\Omega^{-1} + \Sigma^{-1})Z'Z) \right].$$

In order for the computation to be effective, we need to be able to draw from this distribution quickly.

## 4 Discussion

The significance of this distribution and its mixture representations is hopefully clear — the matrix-variate 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 (2.0.1).

## 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.
- Kaitai Fang and Yao-Ting Zhang. *Generalized multivariate analysis*. Springer, 1990.
- Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*, volume 104. CRC Press, 1999.
- Alan Huang and MP Wand. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452, 2013.

Arakaparampil M Mathai. *Jacobians of matrix transformations and functions of matrix argument*. World Scientific, 1997.